Diversity, Distribution and Evolution of the Planktonic Diatom Family Chaetocerotaceae

Thesis

The Open University, London
Stazione Zoologica Anton Dohrn, Naples

Diversity, distribution and evolution of the planktonic diatom family Chaetocerotaceae

Chetan Chandrakant Gaonkar

Thesis submitted for the degree of Doctor in Philosophy (PhD) in Life and Biomolecular Sciences

July 2016

I dedicate this thesis to my mother Savita and my father Chandrakant Gaonkar

Diversity, distribution and evolution of the planktonic diatom family Chaetocerotaceae

Chetan Chandrakant Gaonkar

Abstract

The number and abundance of diatom species in environmental samples are counted traditionally by means of light microscopy (LM). However, recognizing –let alone, counting- species is often challenging because of the existence of cryptic species and intraspecific phenotypic plasticity. Proper characterization requires isolation of cells, growing them into monoclonal cultures, and characterizing the cultures genetically and morphologically. However, not all species grow in culture, featureless ones are less likely to be isolated, and the procedure is laborious. High-throughput sequencing (HTS) metabarcoding bypasses morphology; DNA is collected from environmental samples, a particular marker sequenced, and the resulting sequences sorted into clusters or terminal clades assumed to represent species. Yet, reference barcodes of taxonomically validated species are needed to identify these clades. This exercise is the main aim of my thesis.

Since it is impossible to do this for all the diversity within a PhD thesis project, we selected Chaetocerotaceae, an abundant and diverse family of marine planktonic diatoms, containing two genera: *Chaetoceros* and *Bacteriastrum*. Its members uniquely share setae; thin siliceous tubes emerging from the valve corners, facilitating detection in samples. Strains were obtained from the Gulf of Naples (GoN), from Central Chile and Roscoff – at sites for which LTER time series data are available.

A total of 270 strains were obtained from these sites, and their 18S- and partial 28S rDNA sequences and morphological information gathered. The strains grouped into 60 genetically distinct species, thus providing a dataset of validated Chaetocerotacean 18S reference barcodes. Inferred molecular phylogenies showed monophyletic Chaetocerotaceae as well as monophyletic *Bacteriastrum* inside paraphyletic *Chaetoceros*, and the presence of cryptic diversity. To start with taxonomic updates, the species *C. sporotruncatus* and *C. dichatoensis* were described within the *C. socialis* species-complex based on spore morphology and sequence differences. Several rDNA sequences contained spliceosomal introns (ca. 100bp) and/or group-I introns (ca. 400bp). Phylogenies inferred from the introns did not corroborate rDNA phylogenies, suggesting horizontal gene transfer. Presence/absence of introns in conspecific strains sampled in different seasons suggests population differentiation between these seasons.

A HTS dataset consisting of V4-sequences (part of 18S) from 48 seawater samples taken over the seasons in the GoN revealed 76 terminal clades of which 46 grouped with a reference barcode. Some of these species occur year-round whereas most others are seasonal. Surprisingly, of the 30 clades belonging to unknown Chaetocerotacean species, two appear to be among the most abundant in the GoN.

Acknowledgements

During the three years of my PhD at the Stazione Zoologica Napoli, there are a number of people who guided and provided me support to accomplish this work. Without their help, this work would not have been possible. I would like to express my sincere thanks:

... to my director of studies, Dr. ***Wiebe H.C.F. Kooistra****, for providing me an opportunity to be his student. His guidance and support introduced me to the fascinating world of Phycology.*

... to my internal supervisors Dr. ***Marina Montresor*** *and Dr.* ***Diana Sarno*** *for their patience, ideas, their help and hours of excellent discussions.*

... to my external supervisor Prof. Dr. ***David Mann****, for the fruitful discussions during our meetings. His questions and suggestions have helped me improve my work. I consider it an honour and privilege to have him in my supervising team.*

... to Dr. ***Carina Lange****, for her support and training in handling EM. The time spent with her helped me to learn more things about Chaetoceros identification.*

... to Dr. ***Adriana Zingone*** *for providing the HTS data, the study was supported by the Italian project MIUR-FIRB Biodiversitalia (RBAP10A2T4) and by the MIUR Flagship project RITMARE.*

... to my colleagues and friends Dr. ***Roberta Piredda*** *for introducing me to the world of meta-barcoding; Dr.* ***Maria Paola Tomasino*** *for sharing knowledge and helping in all matter; Dr.* ***Nicolas Maillet*** *for providing me his scripts to solve my complex tasks and for his helpful suggestions over beers.*

... to my friends ***Katerina****,* ***Oriana****,* ***Christian****,* ***Solenn****,* ***Laura V****.,* ***Laura E****.,* ***Maurizio****,* ***Arianna****,* ***Eleonora****,* ***Isabella****,* ***Mariano****,* ***Angela****,* ***Sneha****,* ***Daniele****,* ***Davide****,* ***Gauri, Ravi, Lalita, Rajaneesh, Deepak, Vijay, Sameer, Temjen,*** *and many more whom I met. I appreciate everyone for welcoming me, helping me sort things out in Napoli, sharing thoughts and having fun during numerous occasions to enjoy pizza and drinks together.*

... to the ***MareChaira*** *team for providing plankton samples every week, without which this work would not have been possible. I also like to give my special thanks to te* ***Estación Costera de Investigaciones Marinas*** *(ECIM) and The* ***Department of Oceanography*** *(DOCE) of the* ***University of Concepcion Chile*** *for hosting Dr. Wiebe Kooistra and providing support for collection of the strains from Chile. I thank Dr.* ***Ian Probert****, Station Biologique Roscoff, France, for hosting Dr. Wiebe Kooistra and helping in phytoplankton collection.*

... to ***Ferdinando Tramontano*** *to helping me with my cultures during my study. I thank* ***Alessandro Manfredonia*** *for providing me the culture medium at various occasions. I thank* ***Carmen Minucci*** *for helping me with the molecular biology part.*

... to the ***Molecular Biology Service, SZN*** *for helping me obtaining sequences during my study. I thank the* ***Molecular Biodiversity Lab*** *(MoBiLab) of* ***LifeWatch-Italy*** *(Bari) for sequencing eDNA.*

... to the electron microscopy group, ***Annarita*** *Graziano and* ***Franco Iamunno*** *for their patience and help in my long hours spent at the EM. The collection of my morphological images would not have been possible without their support.*

... to ***Stazione Zoologica Anton Dohrn****,* ***Napoli*** *for accepting me as PhD student, providing financial support to conduct my research. It was my pleasure to be a PhD student in this research institute, which provided me diverse learning opportunities. I thank Gabriella Grossi for her help with all the documents to fill out in time for the OU.*

... finally, I like to thank my parents; ***Savita*** *and* ***Chandrakant Gaonkar*** *for their unconditional love. to my brother* ***Manoj****, thanks for all your love and comforting words. The endless support of my family has meant far more to me than I could possibly express in words.*

List of Tables

List of Figures

List of Figures

BLAST	Basic Local Alignment Search Tool
BI	Bayesian Inference
BPP	Bayesian posterior probability
EM	Electron microscope
GoN	Gulf of Naples
HTS	High Throughput Sequencing
LM	Light Microscopy
LTER-MC	Long-term ecological research station - MareChiara
ML	Maximum Likelihood
MP	Maximum Parsimony
NCBI	National Center for Biotechnology Information
NJ	Neighbour Joining
OTU	Operational Taxonomic Unit
PAUP	Phylogenetic Analysis Using Parsimony
rbcL	Ribulose-Bisphosphate Carboxylase large subunit
SEM	Scanning Electron Microscopy
TEM	Transmission Electron Microscopy

General Introduction

What is phytoplankton; its significance as primary producer

Phytoplankton constitutes the photoautotrophic organisms that occur free drifting in the water column of marine and freshwater environments. It includes a taxonomically diverse and polyphyletic group of organisms, the members of which are classified based on their molecular phylogenetic position and on their types of photosynthetic pigments, storage materials, flagella (if any), cytoplasmic organization and cell coverings into chlorarachniophytes, cryptomonads, cyanobacteria, euglenoids, glaucophytes, haptophytes, dinoflagellates, photosynthetic stramenopiles and green algae (Graham & Wilcox 2000). Arguably, one of the most conspicuous lineages amongst these comprises the diatoms, which form a lineage within the stramenopiles, (Graham & Wilcox 2000).

With ample supply of nutrients light and low densities of adversaries i.e., grazers etc., phytoplankton can increase in biomass rapidly, forming a so-called bloom. Whether a bloom can form and persist or not depends on the intensity of mixing and the mixing depth of the water column (Eppley & Harrison 1975; Reynolds & Walsby 1975). Mixing means that a microalgal cell can find itself at times near the surface, where the summed amount of light received over the daily cycle is high enough for primary production to outpace respiration, and at times at depth, where light is in short supply or altogether absent and respiration outpaces primary production. A bloom can form only primary production outpaces losses due to respiration, grazing, sinking out of the photic zone and other forms of mortality and senescence.

Compensation depth and critical depth are the two critical parameters that determine if a bloom can form or not. The compensation depth is the depth at which photosynthesis equals respiration. Above this depth, biomass increases because growth outruns losses, while, below it, biomass diminishes because losses outrun growth. The critical depth (Sverdrup 1953) defined as, ‘the water column depth at which overall biomass increase due to

photosynthesis equals overall biomass losses.' In other words, the depth at which the integral of net growth rate (photosynthesis minus loss) over the water column above it becomes zero. Photosynthesis decreases from a maximum near the surface to zero at depth. The critical factor that determines whether a bloom can occur or not, is determined by the mixing depth. Mixing intensity depends on the wind, wavelength, and currents, whereas mixing depth depends on the presence of a thermocline or halocline. The critical depth hypothesis (Sverdrup 1953) states that when the mixing depth exceeds the critical depth, loss of biomass exceeds gain and a bloom cannot occur, whereas if the mixing depth is shallower than the shallower than the critical depth, a bloom can occur. Several modifications and refinements of this hypothesis have been proposed, most of which focus on details of bloom dynamics in the spring- and autumn of the temperate zones (Behrenfeld 2010; Chiswell 2011).

Bloom dynamics in the temperate zones

Most studies based on phytoplankton blooms focus on temperate regions during spring and autumn (Gačić *et al.* 2002; Findlay *et al.* 2006). In winter, mixing depth exceeds critical depth because there is no thermocline and the amount of sunlight is low. With the onset of spring, incoming radiation increases pushing the critical depth down. In addition, a thermocline forms, reducing the mixing depth. As soon as the critical depth exceeds the mixing depth, a bloom can take off. Generally, diatoms are the first to bloom because of their rapid growth rates, highly efficient nutrient uptake, and apparent adaptedness to mildly turbulent water and rapidly changing light intensities (Furnas 1990; Savidge *et al.* 1995; Calbet & Landry 2004). Blooms are temporary and deteriorate as soon as conditions become unfavourable, usually upon nutrient depletion.

Diatom cells use silicic acid for building their cell walls, thereby diminishing the concentration of this macronutrient in the water column (Martin- Jézéquel *et al.* 2000). In open oceans, the diatom spring bloom is terminated usually by a shortage of silica. With the

depletion of silicic acid, the phytoplankton community alters from diatom-dominated to flagellate-dominated (Officer & Ryther 1980; Malone *et al.* 1996). Grazing zooplankton rapidly recycles nutrients, but their faecal pellets rain out of the water column, exporting nutrients below the thermocline (Smetacek 1999). During the autumn, the water temperature in the upper layer decreases and turbulence increases due to increased windiness. As a result, the thermocline breaks down and the depth of the mixed layer increases. Nutrients are in plentiful supply again leading to an autumn bloom. Yet, the amount of incoming sunlight diminishes as well, leading to a decrease of the critical depth. Consequently, the autumn bloom terminates as soon as the increasing mixing depth exceeds the diminishing critical depth (Assmy & Smetacek 2012).

With the decline of the favourable conditions, some groups in the phytoplankton can produce dormant stages (Graham & Wilcox 2000) that help to overcome adverse conditions. These resting stages may constitute an indicator of unfavourable environmental conditions or an adaptive response to grazers (Edlund & Stoermer 1993). Resting stages can remain dormant in the sediments for years (French & Hargraves 1980; Gran 1912; McQuoid & Hobson 1996). Because of the substantial deposition of silica in diatom resting spores, they can sink out more rapidly than vegetative cells (French & Hargraves 1980).

Bloom dynamics in other parts of the world

The drivers of bloom formation and termination described above are typical for marine habitats in temperate latitudes. In the Polar region, a phytoplankton bloom can occur only during the summer where light is 24h a day (Mann & Lazier 2006). On the western coasts of the continents, cold currents are usually directed towards the equator. Winds perpendicular upon these currents can cause strong upwelling events during which cold nutrient- and CO_2-laden deep water wells up into the photic zone, giving rise to episodic, but frequent, dense phytoplankton blooms throughout the year. After a couple of days, such

patches of dense blooms are consumed, nutrients rain out in faecal pellets and the bloom peters out until the next upwelling event (Mann & Lazier 2006). A similar type of bloom can be encountered in areas with strong offshore monsoonal winds. The warm oligotrophic water is replaced by cold nutrient-rich water, ensuring a sustained phytoplankton bloom. Nutrients can also be provided by monsoonal nutrient-laden runoff, but in this case, the accompanying silt can hamper phytoplankton growth or promote blooms of smaller phytoplankton because light does not penetrate deep in sediment-laden water (Wiggert & Murtugudde 2007; Hood *et al.* 2009).

Bloom dynamics in the study area, the Gulf of Naples

In the Gulf of Naples (GoN), two types of water masses can be encountered in close proximity: offshore oligotrophic water from the open Tyrrhenian Sea, and coastal, usually more nutrient-rich water. In the offshore waters, the plankton is dominated by minute species occurring in low density, though like in more northern waters spring- and autumn blooms do occur. In contrast, the coastal water always receives nutrients from runoff by small rivers and by the agglomeration of Naples. Therefore, phytoplankton is always present in the water column, density peaks depending on the season and on runoff. Diatoms are common year round but the species composition differs along the seasonal cycle (Scotto di Carlo *et al.* 1985; Ribera d'Alcalà *et al.* 2004: Zingone *et al.* 1995*,* 2010).

Diatoms: General characteristics

The family Chaetocerotaceae belongs to the diatoms (Bacillariophyceae). Diatoms are characterised by a morphologically elaborate compound silica cell wall, called a frustule (see e.g., Round *et al.* 1990). Because of this conspicuous feature, diatoms have been studied extensively ever since the invention of light microscopy (LM)

Morphology

The frustule is composed of two thecae, within which the cell components are organised. Each theca is comprised of a valve and an accompanying set of girdle bands. The valves are shaped like petri dishes or rowing boats, and girdle bands as open hoops. Frustule shape and ornamentation are highly diverse and constitute key characteristics for diatom taxonomy (Round *et al.* 1990).

One set of valve and girdle bands, together, called the 'epitheca,' overlaps a similar set of elements, called the 'hypotheca,' which therefore is minutely narrower or smaller than the epitheca. This arrangement of the frustule facilitates growth and cell division, whilst the diatom cell remains covered and protected by the cell wall elements at all times. Cells grow by increasing their biomass (cell content). Cell wall elements are rigid, but the cell volume increases because the sets of girdle bands can slide along one another in a telescopic fashion and in many species, additional girdle bands can be formed as the cells grow.

Nonetheless, this cell wall architecture poses a constraint. With each mitotic cell division, the epitheca of the mother cell forms a new hypotheca inside it whereas the hypotheca of the mother cell becomes epitheca and forms a new, slightly narrower hypotheca inside it. The result, one daughter cell maintains the same size as the mother cell whereas the other one is minutely narrower. The consequence is that with on-going vegetative growth the average cell size or diameter of a clonal cell line decreases and the variance in cell sizes increases. This is referred as the MacDonald-Pfitzer rule of diatom growth (MacDonald 1869; Pfitzer 1869). Diatom cells can escape from this miniaturization by means of sexual reproduction (see below).

Each cell wall element is composed of high molecular weight proteins organised in a matrix, impregnated with hydrated silica ($SiO_2.nH_2O$). Seawater is rich in dissolved silicic acid. The silica is imported through the plasmalemma and transported through the cytoplasma,

mediated by means of transporter proteins to specialized intracellular compartments, called silica deposition vesicles (SDV), where the silica is precipitated in amorphous form. A gene family encoding the silicic acid transporter (SIT) imports this silicic acid from the environment into the diatom cells (Alverson 2007; Shrestha & Hildebrand 2015). This mechanism of biomineralization makes diatoms key players in the silicon cycling in the world's oceans (Treguer *et al.* 1995).

Depending on the structural plan of the valve symmetry, diatoms are differentiated into pennates and centrics. Pennate diatoms reveal bilateral symmetry, whereas the centric diatoms exhibit radial symmetry. The pennates are subdivided into raphid pennates (those usually possessing a pair of slits, called a raphe, running through the sternum), and araphid pennates (those without such a raphe). The raphid pennates are mostly present in the benthic zone, where the raphe enables them to move actively over surfaces. The silica in the cell wall of a living diatom does not dissolve because, it is protected by an organic coating (Falkowski *et al.* 2004). When the cell perishes, this organic coating is broken down by bacteria and the exposed silica dissolves, unless the frustule elements sink out of the water column rapidly, get covered in the sediments, and become fossilized (Falkowski *et al.* 2004; Pielou 2008).

Ecological importance of diatoms

Diatoms are important players in the global biogeochemical cycles of carbon and silicon and probably contribute ca. 20-25% of the global photosynthetic activity (Field *et al.* 1998; Mann 1999). They form the primary food source for grazing copepods and juvenile fish as well as a range of filter feeders such as molluscs and tunicates. Diatoms are applied as ecological indicators of biotic and abiotic conditions, as well as of anthropogenic and natural impacts (Stevenson *et al.* 1996), mainly because of their high specificity towards environmental parameters, such as nutrients (Borchardt 1996), pH (Smol *et al.* 1986), salinity (Snoeijs

1999), light and temperature (DeNicola 1996; Hill 1996) and hydrological conditions (Dixit *et al.* 1993; Biggs & Hickey 1994; Fritz *et al.* 1999). Fossil frustule elements of vegetative cells and resting spores can be used in paleo-environmental reconstruction (Suto 2006; Stickley *et al.* 2008).

The taxonomic groups of planktonic diatoms

Diatoms can be classified according to their valve shape in radial centrics, bi- and multi-polar centrics, and pennates. The vast majority of diatom species occur in benthic communities where they abound free-living on the sediment or epiphytically, and most of these belong to the pennates. Marine planktonic diatoms are dominated by radial centric-, and bi- and multipolar centric lineages, though a few important pennate groups abound as well. The most diverse and abundant representatives of radial centrics are *Corethron* Castracane, *Coscinodiscus* Ehrenberg, and *Melosira* C. Agardh, which possess labiate processes (rimoportula) along with their valve mantle, and *Leptocylindricus* Cleve and *Rhizosolenia* Brightwell, which lack such processes. Within the bi- and multipolar centrics, the Thalassiosirales Glezer & Makarova and Chaetocerotaceae Ralfs in Pritchard, constitute the most diverse and abundant groups. Thalassiosirales (e.g., *Thalassiosira* Cleve and *Skeletonema* Greville) reveal radial symmetry as well, but this seems to have come about through secondary loss of bi- or multipolar symmetry (Round *et al.* 1990).

Colonies of several groups of bi- and multipolar centrics and some radial centrics (Kooistra *et al*. 2007) are formed by means of inseparable silica connections between the sibling cells (e.g., *Bacteriastrum*, *Chaetoceros, and Skeletonema*). The Class Mediophyceae, a.k.a. bi- and multipolar centric diatoms, constitute a group of species exhibiting polarity in their valves, i.e., their valves are elliptical, elongate, triangular, square, or stellar in outline and show polarity in their ornamentation. Nonetheless, they belong to the centric diatoms because of the striae in the valve face exhibit a radial organization, radiating out from a

roundish pattern centre. Diatom phylogenies inferred from 18S rDNA sequences (Medlin & Kaczmarska 2004; Kaczmarska *et al*. 2007; Kooistra *et al*. 2007) or multiple markers (Alverson *et al*. 2007) indicate that diversification first occurred among the diatoms with a centric organisation, and that pennates evolved later from one of the bipolar centric groups. Thalassiosirales have circular valves, and therefore look superficially like radial centrics, but they exhibit a single large labiate process on the side in the valve mantle. Phylogenies inferred so far suggest that Thalassiosirales group among the polar centric diatoms and Lithodesmiales resolve as the sister group. This indicates that the Thalassiosirales may have lost bi- or multi-polarity secondarily (Alverson *et al*. 2007; Kooistra *et al*. 2007).

In order to prohibit cell chains from becoming intolerably long, certain cells in a string divide without forming such connections, thus allowing the daughter chains to drift apart without damage; see e.g., *Skeletonema* (Sarno *et al.* 2005). Among the pennates, *Pseudo-nitzschia* H. Peragallo in H. Peragallo and Peragallo, *Asterionellopsis* F.E. Round and *Thalassionema* Grunow ex Mereschkowsky can dominate the marine phytoplankton community (Kooistra *et al.* 2007; Not *et al.* 2012). Several *Pseudo-nitzschia* species are potentially harmful because of their potential to produce domoic acid, a toxic amino acid (Lefebvre & Robertson 2010).

Life cycle

Diatoms exhibit a diplontic life cycle, meaning that during most of the life cycle, vegetative diploid cells divide mitotically. Sexual reproduction constitutes just a brief phase in this life cycle; it includes gamete formation, followed immediately by gamete fusion into a zygote from which the initial (diploid) diploid cell emerges.

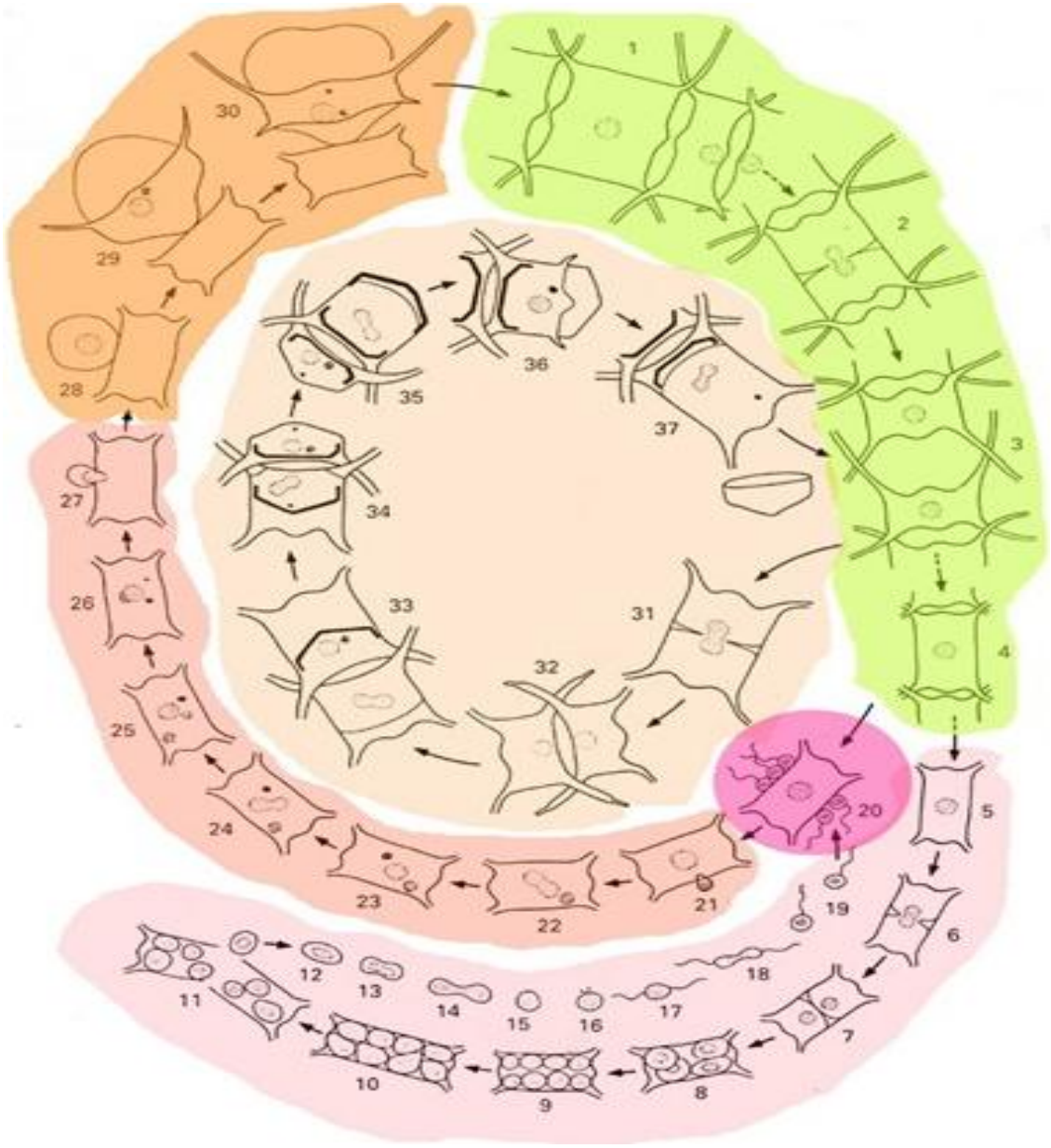


Fig. 1.1: Life cycle of *Chaetoceros* after von Stosch *et al.* (1973) downloaded from http://slideplayer.com/slide/9476226/ on 30/05/2016. The green part illustrates the vegetative phase of the life cycle (clonal reproduction, illustrations 1-4). The pink part (illustrations 5-19) depicts the formation of male flagellated microgametes. Illustration 20 (dark pink) shows the not fully formed female gametangium with microgametes attracted to it. The soft orange part (illustrations 21-27) shows the fusion of the microgamete into the female gametangium whilst the female cell undergoes the first reduction division (illustration 22) and the second one (illustration 24). Illustration 27 shows the moment the haploid nuclei have fused and the cytoplasm starts emerging from the old cell wall along the cingular bands to form the auxospore. The orange part (illustrations 28-30) depicts the auxospore formation. Note the apparent contraction of the protoplasm from the initial auxospore cell wall, in two phases, the first one from the top, and the second one from the bottom of the auxospores initial wall. The initial cell forms in this auxospore to start the vegetative part of the life cycle again. The beige part (illustrations 31-37) depicts the formation (33-35) and germination (36-37) of a resting spore. Note the nuclear divisions without any apparent cell division. The reason is that formation of a new valve is always accompanied by a nuclear division, one nucleus of the formed pair getting lost.

In centric diatoms, sexual reproduction proceeds by means of oogamy, i.e., through the formation of flagellated microgametes and a non-motile macrogamete. In contrast, sexual reproduction in pennate diatoms proceeds by means of isogamy, through the formation of morphologically identical non-flagellated gametes that may or may not show different behaviour prior to gamete fusion. In centric diatoms, a single clonal cell line can form both macrogametes and microgametes (homothallic) whereas in pennates the clonal cell line is either male or female (heterothallic). Upon gamete fusion, the formed zygote swells into a specialized cell, called an auxospore, which re-establishes initial (large) cell size. In some centric diatoms, the newly formed auxospore is encapsulated in an organic cell wall, and the initial vegetative cell breaks out of this confinement. This initial cell is globular, and with subsequent mitotic divisions, the cell assumes a shape like a petridish.

In others, including in *Chaetoceros* and *Bacteriastrum*, the genera in the focus of my thesis, the swelling of the zygote into an auxospore is confined by silica bands, called properizonial bands, which are laid down in sequence and confine the swelling of the zygote in particular directions. In addition, in *Chaetoceros*, protoplasm in the swelling zygote appears to be withdrawn from part of the cell wall (see Fig. 1.1). In these ways, the initial cell obtains a shape different from a globule. Auxospores are not resting spores; auxospores generally possess flimsy silica coverings from which the initial vegetative cell can liberate itself easily. Diatom resting spores, instead, possess frustule elements that are generally more strongly silicified than those of actively growing cells. These resting spores usually develop directly from vegetative cells and can form anytime during the vegetative life cycle (see Round *et al.* 1990). Resting spores are typical for bi- and multipolar centric diatoms; their morphology differs often markedly from that of the normal vegetative cells.

Fossil record

The silica-rich frustules of the vegetative life cycle stage of diatom and their resting spores are the prime units to be preserved in the fossil record. Fossil diatom frustules can be used to address hypotheses about species succession and diversity changes (Anderson & Battarbee 1994). With the help of other paleolimnological techniques, it can be used to reconstruct paleoclimate. Such reconstructions may provide insights in past episodes of rapid climate change. A study conducted by Schmidt *et al.* (1990) reconstructed climate change based on sediment pattern and species composition.

The fossil record and molecular data suggest that diatoms are a relatively young lineage, which appeared during the Mesozoic era, around 250 Mya (Sims *et al.* 2006; Sorhannus 2007). From then on diatoms diversified rapidly. The radial centrics appear to constitute the most ancient lineage, dating back to the early Jurassic period (ca. 190 My). The multipolar centrics appear first in the early Cretaceous, followed by pennates in the Late Cretaceous, and raphid pennates in the Paleocene (ca. 55 My).

Phylogeny of diatoms

Until about 25 years ago, classification and phylogeny of diatoms were solely based on morphological characters (Round *et al.* 1990), and on comparison of fossils and extant species (Sorhannus *et al.* 1995). The Polymerase Chain Reaction (PCR) and modern sequencing techniques have opened new possibilities in identification and phylogeny reconstruction. Molecular approaches focus on the genetic information, stored in DNA-, RNA- or protein marker sequences. Medlin *et al.* (1988, 1991, 1998) were the first to use 18S rDNA (nuclear-encoded SSU rDNA) data to sort out phylogenetic relationships among diatoms.

The ribosomal DNA (rDNA) regions encode for their ribosomal RNAs (rRNA), which form part of the ribosomes; the protein factories of the cell (Woese 1987). The overall secondary structure organisation and functions of these RNAs are basically conserved across eukaryotic and prokaryotic lineages, providing information about phylogenetic relationships among species. The rDNA genes are organised as multiple tandem repeated clusters, called cistrons. Each haploid genome can include up to several hundreds of these cistrons. In Eukaryotes, each cistron possesses genes for 18S, 5.8S, and 28S, interspersed by the non-transcribed spacers ITS1 and ITS2 (internal transcribed spacers); ITS1 is present between the 18S, and 5.8S and ITS2 between 5.8S and 28S gene regions.

The 18S rDNA is considered to evolve slowly compared to other rDNA genes. The 18S gene is undoubtedly one of the most widely used molecular markers for phylogenetic inferences at and above the species-level (Ludwig & Klenk 2001; Alverson 2008). Other reasons behind its wide acceptance as a universal marker are: i) presence in all eukaryotes, ii) rate variation among its different regions, iii) the existence of multiple copies in the genome, iv) conserved function, and v) concerted evolution, which homogenizes all copies within a population (Hillis & Dixon 1991). However, the region is too conserved for inferring phylogenetic relationships among very closely related species or among geographically isolated populations of a single species (Theriot *et al.* 2009; Moniz & Kaczmarska 2010). On contrary, the ca. 1750 base pairs of the 18S rDNA sequences have many potentially variable and therefore possibly phylogenetically informative sites to enable the inference of robust phylogenies, i.e., phylogenies with well-supported clades (Woese 1987).

A huge dataset of 18S rDNA genes is now available (e.g. Ribosomal Database Project like PR2, Guillou *et al.* 2013 and SILVA, Quast *et al.* 2013; Yilmaz *et al.* 2014) comprising thousands of taxonomic reference 18S rDNA sequences, and including references of virtually all diatom genera. In addition, several other genes are investigated for comparative

reasons, and to better resolve phylogenetic relationships. Some of the widely investigated genes are, for instance, the variable part towards the 5'- end of the nuclear-encoded large subunit ribosomal DNA (nuclear-encoded LSU rDNA or 28S rDNA) (Sorhannus *et al.* 1995), the bacterial, plastid and mitochondrial encoded small subunit rDNA gene region (16S rDNA) (Medlin & Kaczmarska 2004), the mitochondrial-encoded cytochrome c oxidase I gene (CO1) (Ehara *et al.* 2000) and the plastid-encoded large subunit of the ribulose-1-5-bisphosphate carboxylase-oxygenase gene (*rbc*L) (Theriot *et al.* 2010), which enhance the resolution of diatom phylogeny. Diatom phylogenies inferred from the SSU or 18S rDNA marker reveal that the centric diatoms are paraphyletic whereas pennates are monophyletic (Kooistra *et al.* 2003; Alverson & Theriot 2005; Sorhannus 2004, 2007; Williams & Kociolek 2007).

The family Chaetocerotaceae

Chaetocerotaceae constitutes an important component of marine plankton communities worldwide, both in abundance and in species diversity (Malviya *et al*. 2016; Piredda *et al*. 2017). Many species are ecologically important because they contribute substantially to phytoplankton blooms in coastal ecosystems, providing nutrition for herbivorous zooplankton (Hendey 1964) and other organisms throughout the marine foodweb, including fish. However, some large species can clog fish gills, and therefore, negatively affect fisheries and fish aquaculture (Yang & Albright 1992; Clément & Lembeye 1993; Fryxell & Hasle 2003).

Classification of Chaetocerotaceae:

Kingdom: Chromista

Phylum: Bacillariophyta

Class: Mediophyceae

Subclass: Chaetocerotophycidae

Order: Chaetocerotales

Family: Chaetocerotaceae Ralfs in Pritchard, 1861

Genus: *Chaetoceros* (Ehrenberg 1844) &

Bacteriastrum (Shadbolt 1854).

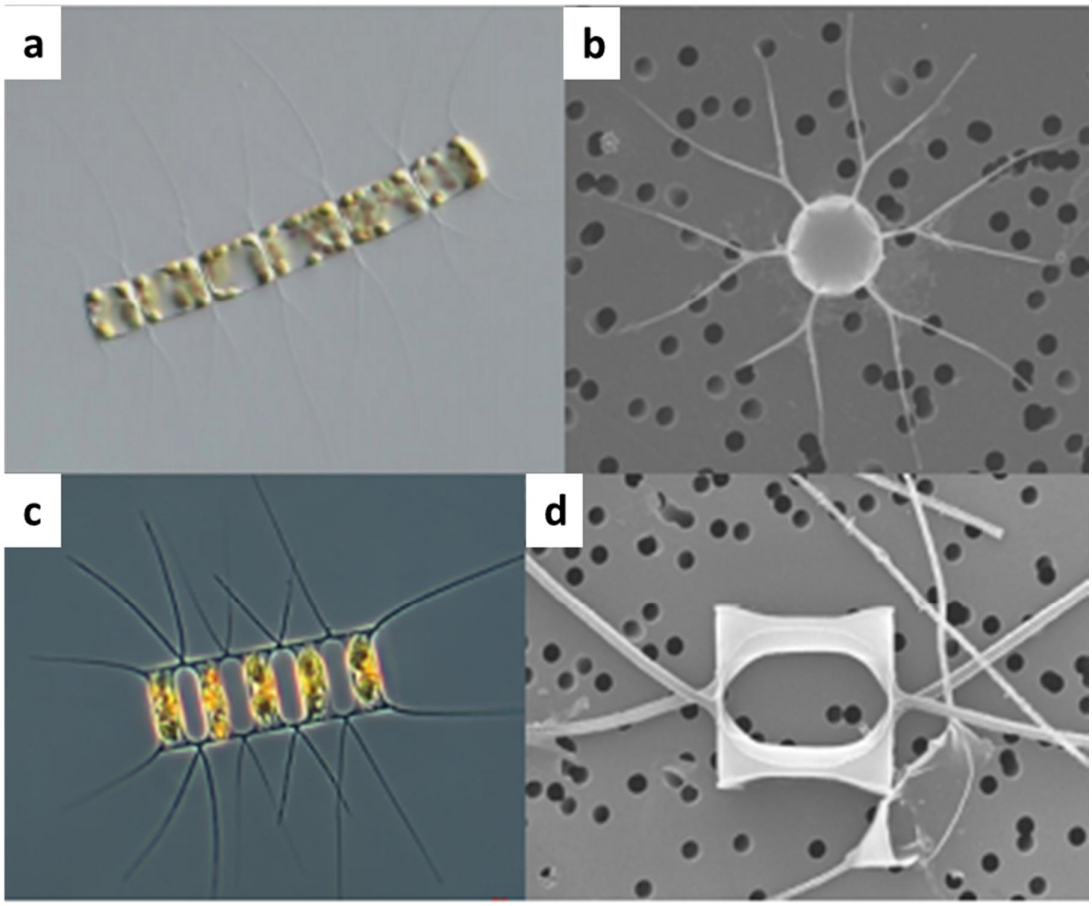


Fig. 1.2: Illustrations of the two genera of the Family Chaetocerotaceae. (a) A vegetative colony of *Bacteriastrum* sp. in girdle view, and (b) an intercalary valve in valve interior view. Note the six pairs of setae fused over a short distance. (c) Vegetative colony of *Chaetoceros decipiens* in valve view, and (d) a pair of intercalary valves and their two pairs of setae, in broad girdle view.

The distinguishing character of the family constitutes the spine-like, tubular projections, called setae, originating from the valve and extending outwards. The setae are formed following the completion of the new hypovalves of the daughter cells. The characteristic features to distinguish species in Chaetocerotaceae are: chain shape and curvature; cell shape; number of plastids per cell; shape, ultrastructure, position, orientation and size of the

setae; size and shape of the space between two adjacent cells in a chain, called aperture; differences between intercalary setae and terminal setae; location and shape of the processes through the valve (rimoportula); the striation and silicification intensity of the frustule; and if present, morphology of the resting spores (Evensen & Hasle 1975; Rines & Hargraves 1988; Hasle & Syvertsen 1997; Rines & Theriot 2003; Kooistra *et al.* 2010). Chaetocerotaceae group within the bipolar centric diatoms (Theriot *et al.* 2010). Specimens are homothallic, i.e. they produce micro and macrogametes in one and the same clonal culture, but in different cells.

In nutrient replete conditions, the cells divide rapidly, forming long chains by setae fusion. Upon nutrient depletion, cells of most species can metamorphose into thick-walled resting spores (Fig. 1.1), which sink to the sea floor, waiting there for favourable conditions to re-commence growth (Durbin 1978; Garrison 1981; Hargraves & French 1983; Eilertsen *et al.* 1995; McQuoid & Hobson 1996; Montresor *et al.* 2013). However, the exact trigger and timing for spore formation and germination is not clearly understood. Photoperiod may be a possible factor triggering spore germination (Eilertsen *et al.*1995).

Characteristics of the genus *Bacteriastrum*

Bacteriastrum is distributed in tropical and temperate zones and is strictly marine (Sarno *et al.* 1997; Bosak *et al.* 2015). Algaebase lists eleven species (Guiry & Guiry 2016). The word '*Bacteriastrum*' is derived from the Greek word bacterion, meaning "small stick," referring to the shape of the chain of closely packed cells, and astrum, meaning "star," which refers to the radiating setae in valve view. Its members reveal multipolar symmetry with a circular outline in valve view, and with fenestrated setae, of which there are six to 20 per valve (Hasle & Syvertsen 1997). Valves possess radial costae originating from a central hyaline area or annulus. *Bacteriastrum furcatum* Shadbolt is considered the type species (holotype) of the genus.

Based on the symmetry of the colony, i.e., on the orientation of the terminal setae, the genus was subdivided into two sections by Pavillard (1925), i.e., *Isomorpha,* which includes species with isomorphic terminal setae, i.e., the two terminal valves in a chain exhibit identically oriented setae, and *Sagittata,* which includes species with dimorphic terminal setae, i.e., the two terminal valves in a chain possess setae that are oriented differently; they are differentiated into anterior and posterior setae (Pavillard 1925). The precise orientation of the terminal setae differs among the species. Species of *Bacteriastrum* usually form chain-like colonies. In most species, intercalary setae cross and fuse with those of their adjacent sister cell for a certain distance and then bifurcate. Obviously, terminal setae do not fuse. Cells are usually quadrangular in girdle view, each containing several chloroplasts. Only terminal valves have a centrally located rimoportula.

Most of the species descriptions available for *Bacteriastrum* are solely based on morphological characters described using light microscopy. Ultrastructural details, observed by electron microscopy, are available for only a limited number of species, including *B. mediterraneum* Pavillard (Bosak *et al.* 2015), *B. furcatum* Shadbolt (Fryxell 1978; Sarno *et al.* 1997; Bosak *et al.* 2015), *B. hyalinum* Lauder (Okuno 1962; Round *et al.* 1990; Kooistra *et al.* 2010; Bosak *et al.* 2015) and *B. jadranum* Godrijan, Marić & Pfannkuchen, emend. Bosak & Sarno (Bosak *et al.* 2012; Godrijan *et al.* 2012; Bosak *et al.* 2015). Molecular phylogenies inferred from partial 28S rDNA sequences revealed that *Bacteriastrum* groups within *Chaetoceros* (Kooistra *et al.* 2010; Bosak *et al.* 2015) supporting findings by Rines & Theriot (2003) who carried out a cladistics analysis of morphological traits.

Characteristics of the genus *Chaetoceros*

The genus *Chaetoceros* is widespread, species-rich, and often abundant in coastal waters and other high-productivity-regions of the world's oceans (Malviya *et al.* 2016). According to Guiry & Guiry (2016), 218 species have been described, most of which are marine (Hasle &

Syvertsen 1997). The word *Chaetoceros* originates from the Greek words 'caith', meaning hair, and 'ceras', meaning horn, referring to the setae. *Chaetoceros* exhibits bipolar symmetry, with valves that are elliptical or lanceolate in valve view and that are adorned with two setae. Cells are usually quadrangular in girdle view with an oval or circular base, possessing one, two, or several chloroplasts. The aperture, i.e., the space between valves of adjacent cells in a chain, varies in size and shape and is delimited by the valve face. Valves possess radial costae originating from a central hyaline or transparent annulus. The rimoportula is a small external opening on the terminal valve and is located either centrally or sub-centrally. Multiple coats of intercalary bands envelop the mantle, protecting the cell content.

A rich but still incompletely known diversity seems to occur in the warm tropical waters (Hernández-Becerril 1996). Species of this family usually bloom in coastal ecosystems, at times with 15 to 20 species occurring simultaneously. Symbiotic relationships are known between the tintinids *Eutintinus* sp. and *C. dadayi* Pavillard and *C. tetrastichon* Cleve and between the ciliate *Vorticella* sp., and *C. coarctatus* Lauder (Hernández-Becerril 1996). In addition, setae of several *Chaetoceros* species are known to host epiphytic organisms such as *Pseudo-nitzschia linea, P. americana* (Lundholm *et al.* 2002) and *Phaeocystis globosa* (Rousseau *et al.* 2007).

Ehrenberg (1844) established the genus *Chaetoceros* and described two species, *C. dichaeta,* and *C. tetrachaeta*. The species *C. dichaeta* is considered the type species of this genus and is one of the dominant species in the plankton of the Southern Ocean (Hernández-Becerril 2002). Traditional classification of this genus is controversial because of inconsistencies in the naming of morphological characters (Rines & Hargraves 1988).

Gran (1897) initiated the classification of the genus *Chaetoceros,* subdividing the genus into two groups i.e. *Chaetoceros* (also known as *Phaeoceros*) and *Hyalochaete*. The subgenus

Chaetoceros is distinguished by the characteristic presence of numerous chloroplasts throughout the cell and in the setae, with robust thick setae and a process on each valve. In contrast, the *Hyalochaete* members are less robust, exhibiting a more fragile cell wall. Its chloroplasts are confined to the main body of the cell; not in the setae (Round *et al.* 1990) and rimoportula located only in the terminal valves.

Ostenfeld (1903) divided these two subgenera into 16 subsections, based on the following morphological characters: the number of chloroplasts, valve morphology, setae orientation, girdle organisation and structure of the resting spore. Hernández-Becerril (1991, 1996, 1998) reorganised the 16 sections by merging a few section and adding a few. As of now, 16 sections are recognized in subgenus *Hyalochaete* and six sections in subgenus *Chaetoceros*. Given the limitation of some morphological characters (Evensen & Hasle 1975; Jensen & Moestrup 1998), given phylogenetic information in disagreement with many of the proposed sections (Kooistra *et al.* 2010), and given the recent discovery of new species, the establishment of a new classification is required.

Genus Chaetoceros Ehrenberg 1844,

Subgenus Chaetoceros Hendey 1964 or Phaeoceros Gran 1897

Section Atlantica Ostenfeld 1903

Section Borealia Ostenfeld 1903

Section Peruviana Hernández-Becerril 1996

Section Coarctata Hernández-Becerril1991

Section Rostrata Hernández-Becerril 1998

Section Tetrastichona Hernández-Becerril1998

Subgenus Hyalochaete Gran 1897

Section Dicladia (Ehrenberg) Gran 1905

Section Cylindrica Ostenfeld 1903

Section Compressa Ostenfeld 1903

Section Protuberantia Ostenfeld 1903

Section Constricta Gran 1897

Section Stenocincta Ostenfeld 1903

Section Laciniosa Ostenfeld 1903

Section Diadema Ostenfeld 1903

Section Diversa Ostenfeld 1903

Section Brevicatenata Gran 1908

Section Curviseta (Ostenfeld) Gran 1903

Section Anastomosantia Ostenfeld 1903

Section Furcilata Ostenfeld 1903

Section Socialia Ostenfeld 1903

Section Simplicia Ostenfeld 1903

Section Conspicua Hernández-Becerril 1993

Subgenus *Bacteriastroidea* Hernández-Becerril 1993

The main aim of this thesis; assessing biodiversity in Chaetocerotaceae

This thesis aims to explore the species diversity in the family Chaetocerotaceae with focus on the Long Term Ecological Research station in the Gulf of Naples. Routine monitoring at this station over several decades has revealed marked seasonal cycles in the species composition, but also the disappearance or appearance of species (Scotto di Carlo *et al.* 1985; Zingone *et al.* 1990; Ribera d'Alcalà *et al.* 2004). So, an accurate inventory is needed to identify possible future changes in the species composition and their seasonal occurrence, given expected changes in the environmental conditions due to global warming. Chaetocerotaceae has been selected because, despite its high diversity and ecological relevance worldwide, only a few groups of species in it have been explored taxonomically using combined morphological and molecular approaches (Kooistra *et al*. 2010; Degerlund

et al. 2012; Chamnansinp *et al*. 2013, 2015). A practical reason for selecting this family is that specimens possess setae, which enables easy detection and isolation of species in plankton samples.

In contrast, accurately identifying specimens down to the species level is, at times, difficult. Many species, even some exceedingly common ones, have been described more than a century ago, based on light microscopic (LM) observations of fixed and cleaned material, and illustrations are often sketchy. At present, details of cells and their cell wall elements are photographed in light and electron microscopy, and molecular tools enable characterisation by means of sequence barcodes. Identifying specimens with these antique descriptions can be challenging, if not impossible. Yet, if they can be linked, the species descriptions need to be amended with information gleaned from all these novel methods.

Morphologically defined Chaetocerotacean species often consist of groups of genetically distinct entities that are difficult, if not impossible, to distinguish in LM (e.g., Kooistra *et al*. 2010; Degerlund *et al*. 2012; Chamnansinp *et al*. 2013, 2015). Identification of such cryptic genetic diversity is important because different genetic entities may represent different species, occurring in different seasons and having different ecological roles. Conversely, morphological differences among specimens do not necessarily imply that these specimens belong to different species; the differences may simply result from plasticity. Morphology can change markedly as cell lines proceed through the vegetative part of their life cycle. In addition, many diatom species show plasticity as a result of environmental factors (e.g. grazers, Pondaven *et al.* 2007; Bergkvist *et al.* 2012; salinity changes, Balzano *et al.* 2011; Sackett *et al.* 2013).

Several Chaetocerotacean cells or chains observed in the field samples do not conform or correspond to any of the known species description in literature (Diana Sarno, pers. Comm.) Such specimens may well belong to species new to science. In fact, papers exploring the

diversity within particular Chaetocerotacean clades generally include newly described species (e.g., Chamnansinp *et al*. 2013, 2015). The more specimens can be identified to the species-level in LM, the more accurate the enumeration of the species in environmental samples can be done.

Many novel techniques are currently being explored and tested to monitor plankton biodiversity in environmental samples on a regular basis without having to culture anything. For example, high-throughput sequencing metabarcode approaches (Sogin *et al.* 2006; Huber *et al.* 2007; Mardis 2008; Amaral-Zettler *et al.* 2009; Shokralla *et al.* 2012) require datasets of accurate reference barcode sequences whereas FlowCam approaches require detailed LM images of accurately identified specimens to train computer programs to rapidly and reliably identify specimens in routine monitoring.

General outline of this thesis, questions addressed and methods explained

The overarching objective of the present study is to investigate the species diversity in the Family Chaetocerotaceae,

Chapter 2 describes the study sites and provides information on strain isolation, their culturing as well as their morphological (LM, SEM, and TEM) and molecular characterisation. The chapter includes a list of the base-substitution models and phylogenetic alghoritms and programs used. Techniques used in single chapters only are presented in the Materials & Methods of the relevant chapters.

The aim of Chapter 3 is to explore the species diversity of this family and to infer 18S and 28S ribosomal RNA gene phylogenies to answer the following questions:

- How well do we know the diversity of Chaetocerotaceae in our study area, the Gulf of Naples?

- How thoroughly have the genera been explored; are there still many new species to be discovered or is our coverage of the group nearly complete?
- How common is cryptic diversity?
- How important is spore morphology versus vegetative morphology in the recognition of new species?
- How do *Bacteriastrum* and *Chaetoceros* relate to each other?

To address these questions, specimens of interest were isolated from environmental samples. Resulting monoclonal strains were photographed using microscopy and barcoded by means of PCR amplification and Sanger sequencing (Sanger *et al.* 1977) and the obtained information compared with taxonomic descriptions to identify the strains. Several different gene regions are used nowadays for the barcoding analysis, including the mitochondrial encoded CO1 (Machida & Knowlton 2012); the plastid encoded *rbc*L (Samanta & Bhadury 2016); the nuclear-encoded D1-D2 region of the 28S (Sonnenberg *et al.* 2007); the 18S rDNA hypervariable regions of V4, V7 and V9 (Crease *et al.* 1998; Amaral-Zettler *et al.* 2009; Malviya *et al.* 2016; Piredda *et al.* 2017). In the present study, the entire 18S and partial 28S rDNA were gathered.

The aim of **Chapter 4** is to assess the species diversity of Chaetocerotaceae in the Gulf of Naples, using a high-throughput sequencing (HTS) metabarcoding approach to answer the following questions.

- Do data from cell counts of Chaetocerotacean species in the plankton samples taken at the long-term ecological research station MareChiara (LTER-MC) agree with the information on species diversity according to HTS data gathered from the same samples?

- How many Chaetocerotacean species are shown in the HTS data, but are without reference barcode?
- How many Chaetocerotacean species are shown in the HTS data that are known from elsewhere but have never been recorded at the MareChiara site?

HTS metabarcoding of a series of environmental samples taken over three years at the LTER MareChiara has generated a huge dataset of V4 metabarcode sequences, which was used to explore the Chaetocerotacean diversity at this station. The 18S rDNA gene is widely used to explore the biodiversity in eukaryotes (Delong & Pace 1991; Stoeck & Epstein 2003; Stoeck *et al.* 2010) and the most commonly used 18S barcode sequence markers include the V9 region (ca. 150bp) and the V4 region (ca. 400bp). The V4 region is favoured above the V9 region because of its larger size and higher phylogenetic resolution, its similar length in all eukaryotes, its high variability of many of its sequence positions (Olsen *et al.* 1986; Woese 1987) and its alignability (Bittner *et al.* 2013). In addition, the highly conserved positions flanking the region allow the design of universal primers. "Taxonomic identification" of all the environmental metabarcodes requires a dataset of properly validated reference sequences (Hebert *et al.* 2003; Tautz *et al.* 2003; Vogler & Monaghan 2007; Pereira *et al.* 2008; Stoeck *et al.* 2010; Pawlowski *et al.* 2011).

Chapter 5 presents a case study in which the aim is to work out in detail the diversity within a complex of genetically closely related species, namely the *C. socialis* complex, and its relatives. A series of strains and the resting spores these strains formed were characterised morphologically, ultra-structurally and genetically in order to answer the following questions:

- How many species can be distinguished within the *C. socialis* complex?
- Are strains morphologically assignable to *C. cinctus* and *C. radicans* also genetically distinct?

- Does spore morphology differ among these species?

To address these questions, the same methodologies were used as described in Chapter 3, but into finer details as the various *C. socialis* like species shared highly similar vegetative morphology.

In the course of the generating reference barcodes in Chapter 3, at times several species failed to amplify. The reason being a large number of 18S and 28S rDNA sequences were found to harbour one to several introns. Therefore, the aim of **Chapter 6** is to characterise these introns and explore their phylogenetic relationships:

- To which groups do these introns belong?
- What are the phylogenetic relationships among the introns within each of the groups?
- Do these phylogenetic relationships of the introns corroborate the phylogenetic relationships of their "host" sequences

Materials and Methods

Study Areas

The study focused mainly on the Long Term Ecological Research station MareChiara (LTER-MC) in the Gulf of Naples (GoN), Tyrrhenian Sea (Mediterranean Sea) because it is sampled weekly and the phytoplankton species composition and associated metadata recorded for every sample. This means that there exists a time-series record of its phytoplankton diversity over several decades. The site shows strong seasonality, resulting in outspoken differences between winter and summer plankton. Nevertheless, Chaetocerotaceae can be found year-round. Having a weekly supply of plankton samples means that cells can be isolated from them and grown into strains throughout the year. Additional *Chaetoceros* strains were collected off the Central Chilean coast during October-November 2013, and in Roscoff (France) in September 2014 (Fig. 2.1). Having strains from these distant sites allowed comparison between the Chaetocerotacean diversity of the warm temperate Mediterranean, a bloom in a Pacific upwelling zone and a bloom in a cooler temperate mega tidal area. These particular sites were also chosen for a practical reason, the presence of marine stations where facilities for sampling and culturing were available.

LTER-MC: The sampling site, LTER-MC, is located two nautical miles offshore from Naples, near the 75 m isobath (40.4°N 14.1°E) at the transition between coastal mesotrophic waters and oligotrophic Tyrrhenian offshore waters. Surface water temperatures vary from 14 ± 1 °C (in March) to 26 ± 1 °C (in August); salinity maxima are 37.9 ± 0.2 psu (in late September-October) and minima are 37.4 ± 0.2 psu (in May). The average surface chlorophyll-a concentration shows a slight increase in winter, which is followed by an annual peak in late spring-summer and again increases in autumn. Temperature, salinity, and chlorophyll exhibit high inter-annual variability. The water column is mixed from December to March and a thermocline builds up in April, leading to a stratified water column with the periodic influx of nutrients from coastal runoff. Stratification weakens over the autumn and

finally collapses in November. Diatoms and nanoflagellates are the most abundant component of the phytoplankton community in terms of cell concentration for most of the year, while diatom and dinoflagellates dominate when considering carbon biomass (Ribera d'Alcalà *et al.* 2004; Scotto di Carlo *et al.* 1985; Zingone *et al.* 1990). This site is sampled on a weekly basis for plankton diversity and physical-chemical parameters.

Sampling sites off the Chilean Coast: Three different sites were sampled for *Chaetoceros* strain isolation off the Chilean coast: Concepción and two sampling sites (El Quisco and S. Antonio) a few miles apart and close to the Biological Station of Las Cruces (Pontificia Universidad Católica de Chile (PUC)).

Concepción: The coastal time series Station 18 (36.3°S 73.1°W, ~90 m water depth) is an upwelling site located ca. 30 km offshore from Concepción in central-south Chile. Sampling is done regularly from 2002 on monthly basis with a range of ecosystem variables by the Centre for Oceanographic Research in the eastern South Pacific (COPAS). The upwelling season occurs during the Austral spring-summer (September-March) and non-upwelling during autumn and winter, whereas heavy rain and river runoff predominate the coastal zone (Faúndez *et al.* 2001; Sobarzo *et al.* 2007). A shallow thermocline is observed during the upwelling season and is weaker and deeper during non-upwelling periods (Montero *et al.* 2007). It is a coastal upwelling ecosystem characterized by high productivity and a broad range of seasonal and interannual variability (Lange & Escribano 2012). Chlorophyll-a concentrations were highest during December and April (Montero *et al.* 2007; Morales & Anabalón 2012). The samples were taken on October 29, 2013.

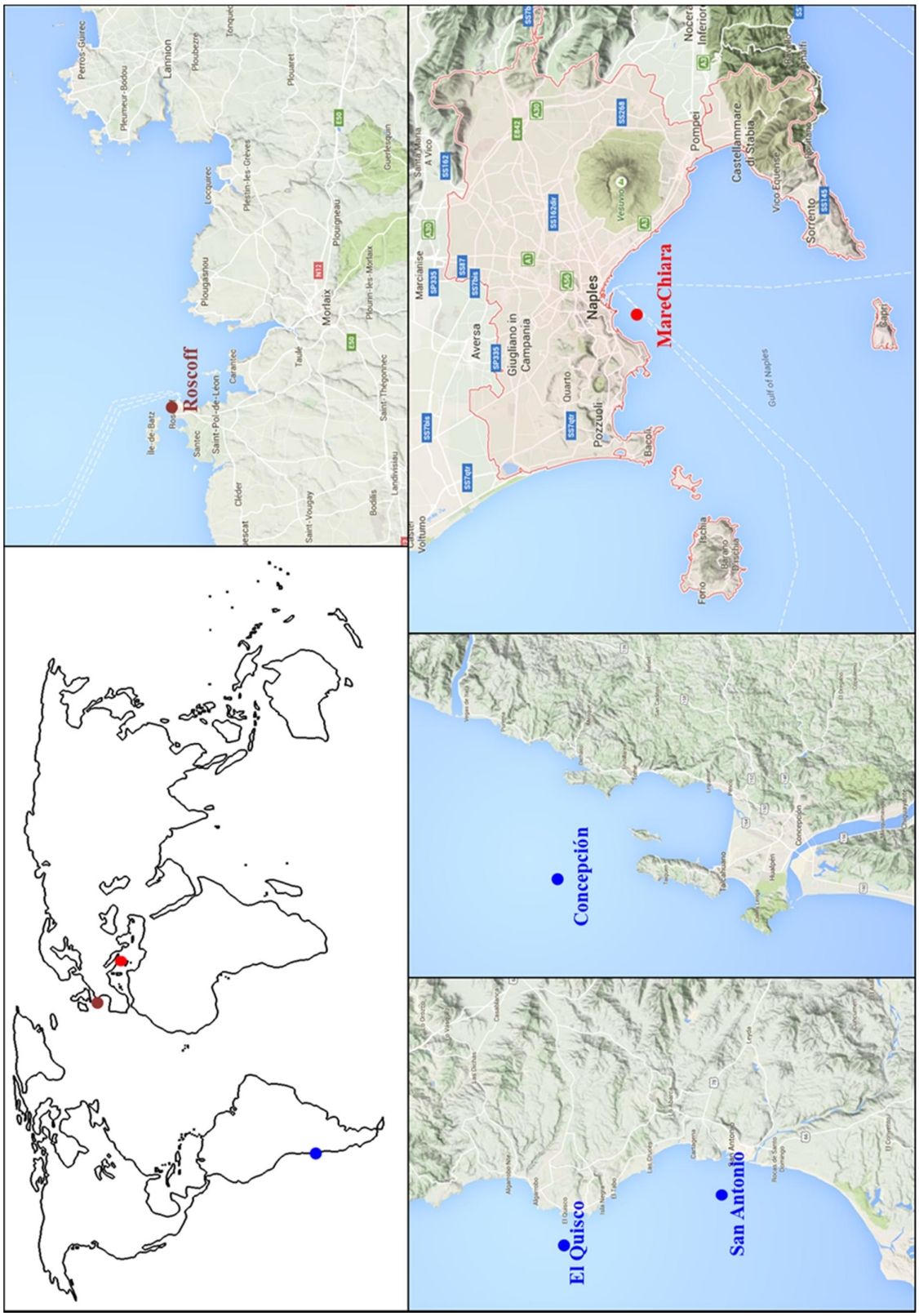


Fig. 2.1: Map showing the sampling sites and location of the area on the world map. Sampling sites: MareChiara (Red) in the Gulf of Naples (Western Mediterranean). Concepcion and Las Cruces (El Quisco and San Antonio; Blue) on the Chilean coast (South Pacific). Estacade, Roscoff (Brown) on the coast of France (North Atlantic).

Las Cruces (LC): The samples were taken ca. 1 mile offshore in the surface water during a strong upwelling bloom (October 16, 2013) and during the demise of the bloom (October 29 and November 4-5, 2013) in El Quisco which is few miles away from LC. The bottom was at this site at ca. 70 m (33.4°S 71.7°W). Studies on coastal dynamics of this region have revealed a strong coastal upwelling with high biological productivity (Avaria *et al.* 1989; Narváez *et al.* 2004). The temperature at LC varies from ca. 12.5 °C during autumn and winter (April-August) and spring and summer (March-July) temperature increases to 16 °C in February (Wieters *et al.* 2003). The salinity at LC is homogenous and ranges between 34.2 psu during July-January with a maxima of 34.5 psu in March (Narváez *et al.* 2004). The average chlorophyll a concentration at the site shows an intermediate pattern of seasonality, with peaks occurring anytime during spring, summer, and autumn (Wieters *et al.* 2003). The samples at the location San Antonio, Chile (33.5°S 71.7°W) were collected during November 1, 2013. The coastal dynamics of this region is similar to nearby LC.

Estacade, Roscoff, France: The sampling site, the tip of the Estacade walking bridge is a coastal site at a depth between zero and 8 m (due to the extreme tides) located in the Bay of Morlaix in the western English Channel (48.4°N 3.6°W). The site is characterised by strong tidal currents and the water column remains mixed throughout the year (Derot *et al.* 2016). The temperature varies from minimum 5 °C in winter to a maximum 18 °C in summer. The sea surface temperature during the collection period was 16 ± 1 °C and was provided by Service d'Observation en Milieu LITtoral (SOMLIT observatory, http://www.sb-roscoff.fr/Somlit/). Salinity at the site varies from 32 – 35.8 psu (Jolivet *et al.* 2015) and chlorophyll content increases from mid-April and reaches a maximum in June, while moderate concentrations occur throughout the late summer and fall (Sournia *et al.* 1987). The sample was collected at high tide.

Phytoplankton sampling: Plankton samples were collected from the surface layer of the water column with a phytoplankton net with a 20 μm mesh size. At the LTER-MC, Dr. Diana Sarno carries out routine phytoplankton identification as follows. The abundance of members of Chaetocerotaceae along with other plankton species is enumerated in Sedgwick-Rafter counting cells using an inverted microscope (Zeiss Axiovert) at 200x magnification. Phytoplankton samples are fixed with neutralized formaldehyde at a final concentration of 0.8%, in order to enable species identification and assess their abundance. Depending on the richness of the samples, 100 ml of the volume is allowed to settle in combined sedimentation chambers, and cells are counted on fields or two transects with an inverted light microscope (Zeiss Axiovert) at 400x magnification using the Utermöhl method (Edler & Elbrächter 2010). Total biomass is calculated from cell numbers times average cell bio-volume of each of the individual species, using geometric formulas (Hillebrand *et al.* 1999; Sun & Liu 2003) and converted to carbon content using the equations suggested for diatoms >3,000 μm^3, and other protists by Menden-Deuer & Lessard (2000). Unidentified *Chaetoceros* species are assigned as *Chaetoceros* species.

Strain collection, cultivation, and characterization: Monoclonal cultures were established by isolating *Chaetoceros* individual cells or chains picked from plankton net samples. Isolation was performed by using fine glass capillaries and an inverted microscope. The isolated strains were purified by washing them in filtered seawater or medium, i.e., by transferring them through a series of clean droplets before putting each of them into their own well in a 12 well tissue culture plate (Costar 3513; Corning Incorporated, NY, USA). The culture medium used was based on f/2 marine enrichment medium (Guillard 1975) and prepared using Guillard's (f/2) Marine Water Enrichment Solution (Sigma-Aldrich St. Louis, USA). Autoclaved seawater, filtered through 0.22 μm pore-diameter filters and with salinity adjusted to 34, was used for medium preparation.

Strains were monitored routinely during the initial days (3 to 7 days) and, when cell density increased, they were transferred to 25-cm^2 (74 ml) polystyrene cell culture flasks (code.430639, Corning Inc., NY, USA) filled with 30–40 ml of f/2+Si medium. Strains were maintained at 20° C, at 12:12 h Light:Dark cycle and an irradiance of 50 µmol photons m^{-2} s^{-1} provided by cool white (40 W) fluorescent tubes. Morphological features of each strain were documented within the first few weeks using light microscopy (LM), as diatoms tend to decrease in cell size and undergo cell deformation after several divisions. LM observation of the strains was carried out on the exponentially growing cultures using a Zeiss Axiophot microscope (Carl Zeiss, Oberkochen, Germany) equipped with Nomarski differential interference contrast, phase contrast, and bright field optics. Micrographs were taken using a Zeiss Axiocam digital camera. An aliquot of the culture was fixed with formaldehyde (final concentration of 4 %) for electron microscopic observation.

Preparation of strains for electron microscopy included a first step in which the fixed material was rinsed several times with distilled water. Upon each washing, the culture was pelleted down at 5000 rpm for 20 min (Eppendorf Centrifuge 5417R). After three washes the pellet was subjected to acid treatment. The ratio of sample to acids was 1:1:2 (Sample: HNO_3:H_2SO_4). This mixture was boiled for a few seconds to remove the organic matter from the valves. On average all samples were boiled for a minute. Thin walled cells were boiled for 30-50 seconds and thick walled cells with organic matter, for 2 min. The sample was washed repeatedly with distilled water until the pH of the material was neutralised. For transmission electron microscopic analysis, the cleaned material was mounted on a Formvar-coated 200 mesh nickel grid and examined with a TEM LEO 912AB (LEO, Oberkochen, Germany). The cleaned material was then mounted on stubs and sputter-coated with gold or platinum. Samples were examined with a JEOL JSM-6500F (JEOL-USA Inc., Peabody, MA, USA) scanning electron microscope.

Strains were identified using the following taxonomic literature: Lauder 1864; Schütt 1895; Cleve 1896; Gran 1897; Karsten 1907; Mangin 1910; Hustedt 1930; Cupp 1943; Corsby & Wood 1958; Hargraves 1972; Evensen & Hasle 1975; Navarro 1982; Rines & Hargraves 1987, 1988, 1990; Giuffrè & Ragusa 1988; Hernández-Becerril 1991, 1992, 1993a, 1993b, 1996, 2010; Marino *et al.* 1991; Sánchez Castillo *et al.* 1992; Ruiz *et al.* 1993; Hasle & Syvertsen 1997; Sarno *et al.* 1997; Hernández-Becerril & Granados 1998; Jensen & Moestrup 1998; Rines 1999; Sar *et al.* 2002; Aké Castillo *et al.* 2004; Ferrario *et al.* 2004; Shevchenko *et al.* 2006; Assmy *et al.* 2008; Sunesen *et al.* 2008; Hernández-Becerril *et al.* 2010; Kooistra *et al.* 2010; Lee & Lee 2011, 2014; Bosak *et al.* 2012, 2015; Kownacka *et al.* 2013; Li *et al.* 2013, 2015; Lee *et al.* 2014a, 2014b; Fuad *et al.* 2015.

Spore induction: Resting spores are found in most *Chaetoceros* species and provide valuable information for species identification. Resting spores are formed in order to escape adverse conditions in nature, i.e. darkness, temperature, nutrient depletion, grazing etc. Most studies state the importance of nutrients (specifically Nitrogen) in resting spore formation (Kuwata *et al.* 1993; Sugie & Kuma 2008). In this study, induce resting spore formation was tried in some species of *Chaetoceros*, using different types of medium;

Nutrient-depleted medium (f/2+Si-N): the medium comprised of normal Guillard's medium with silica, but lacking any source of nitrogen. Exponentially growing cells (few chains) were inoculated in this medium and incubated at 15 °C using fluorescent light of 40 µmol photons m^{-2} s^{-1} intensity with 12h Light: 12h Dark cycle. Subjected strains were observed regularly in LM for five weeks.

Nutrient-depleted medium with blue light: few chains of *Chaetoceros* species were inoculated in nitrogen-depleted medium (f/2 medium and f/10 medium), and incubated in ambient light on a north-facing window (no direct sunlight) at 20 °C for the first week. These strains were then transferred the cultures to another chamber with blue irradiance at 20 °C

and incubated them for four weeks. Observation on the culture status and spores were conducted every two days.

T-medium: T-medium (Guillard & Ryther 1962) contains ammonia at 15 µM, phosphate at 7 µM, silicate at 50 µM, trace metals; vitamins, and Trizma. Exponentially growing cells (few chains) were inoculated in this medium and incubated at 15 °C using the fluorescent light of 40 µmol photons m^{-2} s^{-1} intensity with 12h Light: 12h Dark cycle. Subjected strains were observed periodically in LM for five weeks.

Biological factors: During this study, the the possibility that spore formation is induced by the presence of biological agents (such as viruses or bacteria) was explored. Fresh sea water was collected from the Gulf of Naples and filtered through a 5 µm mesh. The resulting filtered solution was diluted with f/2 medium in 1:1 ratio (vol:vol). A few chains of *Chaetoceros* cells were inoculated into this enriched seawater medium. Observations on the culture status and spores were conducted periodically.

DNA Extraction: All isolated strains were subjected to genomic DNA extraction using a CTAB DNA extraction protocol (modified from Doyle & Doyle 1987; Cullings 1992). A small aliquot of the culture was drawn from the bottom of the culture flask with a Pasteur pipette and dispensed in a 1.5 ml sterile micro Eppendorf tube. The suspended cells were centrifuged at 12,000 rpm for 5 min to obtain a pellet. The supernatant was discarded without disturbing the pellet. To this cell pellet, 500 µl of CTAB Lysis buffer (AppliChem GmbH, Darmstadt, Germany) and 12 µl of β-mercaptoethanol (Sigma-Aldrich, Missouri, USA) were added under a fume hood, and the content of the tube was thoroughly suspended by vigorous mixing. This homogenous mixture was then incubated in a pre-heated water bath at 65 °C for 1hr; followed by vortexing for few seconds at 15 min intervals. After incubation, the suspension was placed in an ice bath for 5 min and 500 µl of SEVAG solution (chloroform: isoamyl alcohol at 24:1 vol:vol) was added to this cooled mixture, and mixed. This solution

was then centrifuged at 14,000 rpm for 30 min at 4 °C, which sorted the mixture into three layers. The top aqueous layer contains nucleic acids and water-soluble carbohydrates, the middle layer comprises proteins and other cell debris, and the bottom layer, the chloroform with all the fat-soluble cell components such as chlorophyll. Only the upper aqueous layer was collected in a new tube (ca. 400 µl), without disturbing the middle layer. To this aqueous solution, 400 µl of ice-cold isopropanol was added and gently mixed by inverting the tube upside down; the tube was then incubated at -20 °C for at least 1 hr or at 4 °C overnight. The DNA was pelleted by centrifuging the mixture at 14,000 rpm for 30 min at 4 °C. The obtained DNA pellet was washed with 400 µl of 75% ice-cold ethanol to remove any salts, and centrifuged at 14,000 rpm for 15 min at 4 °C. Residual ethanol was removed by drying the DNA pellet under a fume hood. The obtained DNA pellet was then suspended in sterile milliQ water or TE buffer and stored at -20 °C.

Polymerase Chain Reaction (PCR): For the phylogenetic- and barcode studies, I used two marker gene regions from the nuclear encoded ribosomal subunit (rDNA). The D1 - D3 region - a product of ca 750 bp - of the nuclear encoded large subunit (LSU or 28S rDNA), was amplified. The marker gene was PCR-amplified in 25 µl-volumes containing 10-250 ng DNA, 1 mM dNTPs, 0.5 µM of the D1R forward primer, 0.5 µM of the D3Ca reverse primer, 1x Roche diagnostics PCR reaction buffer (Roche Diagnostics, GmbH, Mannheim, Germany), and 1 unit Taq DNA Polymerase (Roche). The thermocycler (C1000 Touch, BioRad, California, USA) was preheated at 98 °C prior to the PCR cycling. The PCR cycling comprised an initial 3 min heating step at 94 °C, followed by 35 cycles of 94 °C for 35 s, 54 °C for 35 s, and 72 °C for 2 min, tailed by a final extension at 72 °C for 15 min with a final hold at 12 °C.

The nuclear encoded small subunit (SSU) ribosomal DNA (or 18S rDNA; normally ca. 1,800 bp long) was amplified usually with the standard SSU-F/SSU-R primers (Table 2.1). The

PCR mixture was prepared in a similar way as in the 28S PCR mix, in which only the primers were different. The PCR cycles were carried out at an initial denaturation step of 3 min at 94 °C, followed by 40 cycles of denaturation steps at 94 °C for 35 sec, annealing step at 52 °C for 35 sec and an extension step at 72 °C for 3 min, following another extension step of 10 min at 72 °C and a final hold at 12 °C.

Table 2.1. List of Primers used for amplification as well as sequencing. Nucleotides indicated in red are the additional modification specifically done for Chaetocerotaceae, in comparison to the actual primers indicated by reference. Nucleotides in blue denote ambiguity in the primer to accommodate variation in the primer targets in the Chaetocerotacean sequences.

Primer Name	Sequence (5'→3')	Reference
D1R	ACCCGCTGAATTTAAGCATA	Scholin *et al.* 1994
D2C2	CCTTGGTCCGTGTTTCAAGA	Scholin *et al.* 1994
D2Ra	TGAAAAGGACTTTGAAAAGA	Scholin *et al.* 1994
SSU-F	TCYAAGGAAGGCAGCAGGCGC	Hamsher *et al.* 2011
SSU-R	GTTTCAGCCTTGCGACCATACTCC	Ki *et al.* 2007
Ch-11F	TGATCCTGCCAGTAGTCATACGCT	Alverson *et al.* 2007
Ch-82F	TTGAAACTGCGAAYGGCTCAT	Modified from Kühn *et al.* 2006
Ch-300F	ATTAGGGTTTGATTCCGGAGAGG	This study
Ch-528F	GCGGTAATTCCAGCTCCAATAGCGT	Modified from Kühn *et al.* 2006
Ch-536R	AGCTCCAATTACCGCGGCTGCTGGCA	Modified from Kühn *et al.* 2006
Ch-690F	TMAGAGGTGAAATTCTTAG	This study
Ch-690R	TCTAAGAATTTCACCTCTKA	This study
Ch-1055R	TAAGAACGGCCATGCACCACCACC	Modified from Kühn *et al.* 2006
Ch-1055F	GTGGTGGTGCATGGCCGTTCTTAG	Modified from Kühn *et al.* 2006
1147R	AGTTTCAGCCTTGCGACCATAC	Alverson *et al.* 2007
Ch-1147F	GGTCGCAAGGCTGAAACT	This study
Ch-1400R	ACGGGCGGTGTGTACAAAGGGCA	Modified from Kühn *et al.* 2006
Ch-1400F	CCTTTGTACACACCGCCCGTCGCA	This study
TAR-EukF1	CCAGCAGCCGCGGTAATTCC	Stoeck *et al.* 2010
TAR-EukR	ACTTTCGTTCTTGATYAATGA	Modified from Stoeck *et al.* 2010
Ch-V9R	CCTTCYGCAGGTTCACCTAC	Amaral-Zettler *et al.* 2009

Quantity and length of PCR products were examined by agarose gel electrophoresis against known standards. A total of 1 µL of the PCR products was visualized using 1% agarose gel electrophoresis with TBE buffer against 100 bp standard ladders. PCR products showing a single band were purified using a MicroCLEAN kit (Microzone Ltd., Haywards Heath, UK)

following manufacturer's instructions. In the case of multiple bands, the remaining PCR product was loaded on an agarose gel and the target band was excised under low UV light, and subsequent purification of the desired fragment was carried out using DNA Isolation Spin Kit Agarose (PanReac Applichem GmbH, Darmstadt, Germany) following manufacturer's instructions. In the case of failure to obtain PCR product, internal primers were used to obtain a set of products comprising overlapping regions of the 18S region. Short fragments of the 18S regions were amplified and sequenced to obtain partial sequences.

Phusion DNA Polymerase and Troubleshooting: In several Chaetocerotacean strains, the normal Roche Taq polymerase failed to produce any PCR product. Whenever possible, several strains were processed for each species, and it was found that these failures were generally taxon-related. In such cases, the entire sequence was obtained not in one single PCR product, but in two or three overlapping ones, generated with combinations of external and internal primers. Obtained fragments were far longer than expected because of the presence of one or several introns, making the entire sequence up to over 2000 bp longer. If that failed, too, a high fidelity Phusion DNA polymerase (New England BioLabs Inc, Massachusetts, USA) was used to obtain PCR products. This enzyme enhances the fidelity compared to the normal Taq polymerase (Frey & Suppman 1995). The PCR mix was enriched with GC-rich buffer, as it is known to improve the performance of Phusion DNA polymerase on long templates, GC-rich templates, and on those with a complex secondary structure (Frey & Suppman 1995). The Phusion PCR mix of 25 µl contained <250 ng of genomic DNA, 200 mM of dNTPs, 0.5 µM of forward primer, 0.5 µM of reverse primer, 1x of 5X Phusion GC Buffer, and 1 unit Taq DNA Polymerase. DMSO (3%) was added to the PCR mixture to facilitate strand separation of GC-rich templates by base pair disruption, increasing yield and specificity. The PCR cycles were carried out with an initial denaturation step of 30 seconds at 98 °C, followed by 35 cycles of a denaturation step at 98 °C for 10 sec, an annealing step at 58 °C (for 28S) or 60 °C (for 18S) for 30 sec and an extension step at

72 °C for 1 min. Then followed a final extension step of 10 min at 72 °C and a hold at 12 °C until further handling. In the case of failure of amplification with this enzyme, no further attempts were undertaken.

Sequencing of PCR products: For 28S sequencing, the primers D1R, D2C2, and D3Ca were used to obtain the sequence (Table 2.1). The internal reverse primer D2C2 was used in case inserts were present. The sequencing mixture consisted of 15 fmol·µl^{-1} of the PCR product and 4.5 fmol·µl^{-1} of either the forward or reverse primer to a final volume of 10 µl. In the case of SSU gene region, PCR primers along with internal primers were used to obtain a full-length sequence (Table 2.1). The PCR products were sequenced by the Molecular Biology Services at SZN using the BigDye Terminator Cycle Sequencing technology (Applied Biosystems, Foster City, CA), purified using a 'Biomek FX' (Beckman Coulter, Fullerton, CA, USA) robotic station, and analyzed on an Automated Capillary Electrophoresis Sequencer '3730 DNA Analyzer' (Applied Biosystems, CA, USA). Forward and reverse sequences were combined into contigs in BioEdit (Ibis Biosciences, Carlsbad CA, USA). The sequences were manually processed using BioEdit v7.0.0 (Hall 1999). Any site showing an ambiguity in the forward and reverse sequence was recorded as such if the surrounding sites read without any difficulties.

Sequence analysis and alignment dataset: 276 strains of *Chaetoceros* and *Bacteriastrum* were processed to generate the Chaetocerotaceae 28S and 18S reference dataset, and used to infer phylogenies. The LSU alignment included a total of 417 sequences, generated in my study or available from GenBank. The 18S alignment comprised 262 sequences from family Chaetocerotaceae. Additional GenBank sequences belonging to Chaetocerotacean species for which the morphology has been documented, or belonging to unidentified Chaetocerotacean clones, as well as environmental Chaetocerotacean sequences from known geographic provenance were downloaded from GenBank and aligned using MAFFT v7.245

(Katho *et al.* 2002) under default settings and output-in-alignment-order. The alignment obtained was customized manually for minor adjustments in the marker regions using the alignment program SeaView v4.5.4 (Gouy *et al.* 2010).

Phylogeny: In the present study, phylogenies for each data set were inferred using neighbor-joining, maximum parsimony, maximum likelihood, and Bayesian Inference.

Neighbour joining (NJ): NJ is a distance based method for reconstructing phylogenetic trees that use pairwise distances to compute all the branch lengths of the tree. The NJ analysis is fast, which makes it practical for analysing large datasets and for bootstrapping, though the disadvantage is that its results tend to deviate from the true outcome if substitution rates differ among lineages and if there are long branches in the tree. In the latter case, such long branches attract each other because base positions that are identical as a result of homoplasy are simply treated as identical because of common descent. In the present analysis, Kimura 2-parameter (K2P)-distances was used. This model makes a distinction between two types of substitutions in the sequence, assuming that transitions (A↔G and C↔T) are more likely than transversions (A↔C, A↔T, C↔G and G↔T). Distance based NJ trees were generated using K2P model (1000 replicates).

Maximum parsimony (MP): MP is a character-based method that infers a phylogenetic tree that requires the least number of changes along its branches. MP is a cladistics approach, based on shared derived character states (synapomorphies). Although it is not a distance approach, the branch length of an internode in a tree is equal to the number of inferred base changes along that internode (synapomorphies, symplesiomorphies, and homoplasies). The MP analyses were done using PAUP* v4.0a146 (Swofford 2003). All parsimony trees were inferred using the heuristic search algorithm, with ten step-wise additions of the sequences in random order, and tree-bisection-reconnection (TBR) branch swapping to find the most parsimonious tree topologies. Character state optimization was carried out using accelerated

transformation (ACCTRAN, to place changes towards the base of the tree rather than towards the tips, if equally parsimonious). Multistate bases were considered as polymorphisms and gaps were treated as missing data. Bootstrap support values for the internal internodes were obtained based on 1000 bootstrap replicates with one random stepwise addition of the sequences for each bootstrap replicate.

Maximum likelihood (ML): ML is also a character based approach used for phylogenetic inference. Maximum likelihood (ML) finds a tree that maximizes the likelihood of the sequence data, given a particular tree topology and a model of sequence evolution. The ML method searches for a phylogenetic tree with the highest probability or likelihood. A General-Time-Reversible (GTR) model of base substitution was estimated from the dataset. I also incorporate the gamma distribution for rate heterogeneity among sites. ML analyses were performed using raxmlGUI v1.5beta (Silvestro & Michalak 2012). ML trees were generated in ten replicate runs using the rapid hill climbing mode. Support for internal internodes was estimated using bootstrap analysis (1000 replicates).

Bayesian inference (BI): BI is a statistical procedure to estimate parameters of an underlying distribution based on the observed distribution and tries to find a tree that maximises the posterior probability of the data and a model of evolution. Bayesian trees were constructed using MrBayes 3.2.2 on XSEDE (Ronquist & Huelsenbeck 2003) with GTR + r + PINVAR parameters being estimated during the run and using the default value of four Markov chains. The ‘‘temperature’’ parameter was set to 0.2, resulting in incremental heating of each chain. The Monte Carlo Markov Chain (MCMC) length was set to 1 million generations, with a posterior probability of bipartitions (i.e. split frequencies) sampled every 100 generations and diagnosed every 1000 generations. For analysis, I remove the initial 25% of the sampled trees as burn-in. The BI-consensus trees were generated from the sampled trees.

ML tree inference using FastTree: FastTree (Price *et al.* 2009, 2010) uses several stages to generate an optimal tree. It first uses a heuristic variant of NJ to obtain a rough topology by hill-climbing. In the next step, it reduces the length of the tree using a mix of nearest-neighbor interchanges (NNIs) and subtree-prune-regraft moves (SPRs). Following this, it improves the tree – both the topology and branch length – with ML rearrangements. FastTree uses GTR models of nucleotide substitution and accounts for variable rates of evolution across sites (gamma). Later, it estimates the reliability of each split in the tree using the Shimodaira-Hasegawa (SH) test for all the topologies (NNIs). FastTree uses 1000 resamples and does not reoptimise the branch length for the resampled alignments.

Species identification from the database: In order to quickly identify the sequence generated and to assign it a position in the alignment, a local database of 28S and 18S sequences was generated, using the BLAST (NCBI C toolkit) software. The 28S dataset was generated using 'formatdb' command for and is as follows:

"*formatdb -i Chaetoceros_28S.fasta -p F -o T -n 28S_Database*"

where, -i is the Input file (unaligned), either in .fasta or in .txt format

-p is the Type of the file (T for protein and F for nucleotide sequence),

-o is Parse options with output

T - True: Parse SeqId and create indexes or

F - False: Do not parse SeqId. Do not create indexes.

-n is a Base name for BLAST files.

Formatdb is a now discontinued software from the NCBI used in molecular bioinformatics to format protein or nucleotide databases for BLAST, now replaced by makeblastdb. This process generated a database that contains reference sequences from all Chaetocerotacean species. A separate database was generated for the 18S sequences of Chaetocerotaceae.

In order to find which sequence is the best hit for the query sequence with our known database, the following command was used:

"*blastall.exe –p blastn -d 28S_Database –i query.fasta –o Result*"

where, -i is the Input file (unaligned), either in .fasta or in .txt format,

-p is Type of file (default F),

-d is the dataset,

-o is output.

The BLAST program recognizes the query sequence in the dataset and generates a matrices file that provides the best match to the reference sequences with percentage, length coverage, E-value. This file corresponds similarly to the BLAST search generated by the NCBI nucleotide blast online program (blastn). We initiated this step due to the scarcity of reference sequences of Chaetocerotaceae in the GenBank. Also, this step is a quicker means to identify a specimen and assign a position for the sequence in the large alignment before inferring the phylogeny.

Phylogeny of the marine planktonic diatom family Chaetocerotaceae

Abstract

Members of the family Chaetocerotaceae are major constituents in the marine phytoplankton. The main objective of the present study was to assess the morphological and genetic diversity within the family. Therefore, 270 strains were collected from coastal marine habitats at Las Cruces and Concepcion, Central Chile, Roscoff, France and from the LTER-MC, Naples, Italy. All the isolated strains were characterized morphologically and genetically into 60 species. Phylogenies were reconstructed from partial 28S and complete 18S rDNA sequences, using maximum parsimony, maximum likelihood, and Bayesian inference, to assess relationships among the various species. The phylogenies revealed that *Bacteriastrum* forms a clade within paraphyletic *Chaetoceros*. Moreover, the early ramifications in both trees show weak support if any at all. In the partial 28S rDNA phylogeny, the clade containing *Eucampia* resolved inside *Chaetoceros* whereas, in the 18S rDNA tree, it was recovered outside *Chaetoceros,* grouping with other close relatives of *Chaetoceros*. The relationships among the Chaetocerotaceae are not well resolved with the nuclear rDNA, but at least the tree topologies corroborate the single acquisition of the setae.

An additional objective of this study was to obtain barcodes for all the genetically distinct species assessed in this study, including species new to science. Such reference barcodes enable taxonomic identification of terminal clades in phylogenies inferred from High-Throughput Sequencing (HTS) metabarcode haplotypes generated from environmental DNA samples. Since previous studies conducted in our laboratory have chosen the hypervariable V4 region in the 18S rDNA as a metabarcode region and have gathered an environmental metabarcode dataset from 48 surface seawater samples at the LTER-MC site over the years 2011-2013, we use this dataset to assess diversity and seasonality of Chaetocerotaceae. A check of the V4-primers against the obtained 18S rDNA data showed that metabarcodes of all but a few species never reported at the LTER can be PCR-amplified, i.e., should be detectable. Moreover, all entities differing in their 28S and 18S sequences from other such entities also differed in their V4 metabarcode, i.e., can be identified. Finally yet importantly, phylogenies inferred from the V4 sequences revealed a topology similar to those inferred from the entire 18S rDNA sequences, in which Chaetocerotacean sequences resolve in a clade. This enables the gathering of all and only Chaetocerotacean metabarcodes from the abovementioned environmental ones, without having to deal with huge numbers of false-positives and false-negatives.

Introduction

Chaetocerotaceae Ralfs in Pritchard 1861 is one of the most species-rich diatom families in the marine phytoplankton (Guiry & Guiry 2016). Its members are easily recognized because they possess hollow siliceous tubular structures, called setae, emerging from the corners of the valves. Species can be distinguished based on a series of morphological and ultrastructural characters associated with the vegetative cells as well as with the morphologically elaborate resting spores that most species form. The family includes two genera: *Chaetoceros* Ehrenberg and *Bacteriastrum* Shadbolt (Round *et al.* 1990). *Chaetoceros* exhibits bilateral symmetry (bipolar) and possesses two setae per valve while *Bacteriastrum* shows radial (multipolar) symmetry and possesses more than two setae per valve. As per the recent update on the Algaebase (Guiry & Guiry 2016), *Chaetoceros* is morphologically and taxonomically diverse, comprising over 218 species, which are validated based on appropriate taxonomic literature, mostly marine. *Bacteriastrum* includes only 15 such species, all strictly marine. Gran subdivided the genus *Chaetoceros* into two subgenera: *Chaetoceros* Hendey 1964 (formerly *Phaeoceros* Gran 1897) and *Hyalochaete* Gran 1897. This taxonomic distinction was based on whether chloroplasts occur in the setae (*Chaetoceros*) or not (*Hyalochaete*).

The overarching objective of the present study was to know the diversity of Chaetocerotaceae in the study area, the Gulf of Naples, by exploring the morphological, ultrastructural, and genetic diversity. Therefore, cells or colonies (chains) were isolated from plankton net samples, grown into monoclonal cultures (strains), and characterised using LM and EM as well as PCR and Sanger sequencing of two marker regions. The partial 28S rRNA-encoding region (28S rDNA) was acquired because previous studies of the family used this marker (Kooistra *et al.* 2010; Degerlund *et al.* 2012; Chanmansinp *et al.* 2013, 2015; Li *et al.* 2013, 2015; Bosak *et al.* 2015). The full-length 18S rRNA-encoding region (18S rDNA) was obtained as well because it is universally applied in taxonomic studies, and large datasets of reference sequences exist

for most Eukaryotes (Guillou *et al.* 2013; Quast *et al.* 2013); though not yet for Chaetocerotaceae. I focused my study on the Chaetocerotacean diversity at the Long Term Ecological Research (LTER) station MareChiara in the Gulf of Naples (GoN; Tyrrhenian Sea), because weekly monitoring at this site has uncovered a large number of species and marked seasonality in their occurrence (Scotto di Carlo *et al.* 1985; Zingone *et al.* 1990, 2010; Zingone & Sarno 2001; Ribera d'Alcalà *et al.* 2004). The monitoring also revealed several Chaetocerotacean entities not fitting any species description, as well as groups of similar entities that each may represent taxonomically distinct species.

Moreover, several common species such as *Chaetoceros socialis*, *C. tenussimus* and *C. curvisetus* occur year-round under a remarkably wide range of environmental conditions and might constitute complexes of multiple species (see examples in Kooistra *et al.* 2010). This study also focuses on revealing the cryptic diversity (see Chapter 5). The material was collected also from an upwelling zone in central Chile and a mega-tidal habitat in Brittany, France, to compare collected strains from all these ecologically and geographically different places, to assess if the taxa found there are detectable in the Neapolitan diversity as well and if geographical variation exists within morphologically perceived species. Molecular phylogenies were inferred from the two marker regions to reconstruct the acquisition history of morphological and ultrastructural, to assess the relationship between *Bacteriastrum* and *Chaetoceros* and, if *Bacteriastrum* forms a clade inside *Chaetoceros* as in Rines & Theriot (2003) and in Kooistra *et al.* (2010) and if the subgenera *Hyalochaete* and *Chaetoceros* are monophyletic. In addition, incorporation of environmental sequences of unknown species in the phylogenies permits the establishment of hypotheses about their morphology, which enables informed searches for specimens in field samples.

An additional objective was to establish a taxonomically validated Chaetocerotacean reference barcode dataset to enable identification of Chaetocerotacean in the high-throughput sequencing

metabarcode haplotypes generated from plankton samples gathered at the LTER MareChiara. The hypervariable V4-region (Hu *et al.* 2015; Massana *et al.* 2015) of the 18S rDNA has been chosen as a barcode in many previous studies (see Chapter 4) and a V4-dataset actually already exists for the MareChiara site. In the present study, I used the gathered Chaetocerotacean 18S sequences to assess if the universal V4-primers amplify the target regions of all taxa across the family, if the V4-region is variable enough to distinguish among all the uncovered Chaetocerotacean species, and if the Chaetocerotacean V4 sequences form a clade with an internal topology similar to that in the 18S tree (to minimise exclusion of false negative metabarcodes and inclusion of false positives).

Materials and Methods

Strain isolation, cultivation, and characterisation: descriptions of strain isolation, cultivation and characterisation have been detailed in Chapter 2. Strains were documented and identified using taxonomic literature (Table 3.1).

Table 3.1. List of Chaetocerotacean strains and diatom outgroups used in the study. The strains isolated in this study are marked in bold. Some species have a provisional name enclosed in quotation marks. *** = GenBank code of the sequences produced in this study (to be deposited). # indicates the sequence used in the phylogenetic analysis.

Strain name	Strain Code	Collection site [1]	Collection date	Images	Reference for morphological identification [2]	GenBank 28S rDNA	GenBank 18S rDNA
B. cf. *furcatum*	**Na8A3#**	GoN	Feb. 6, 2014	LM, SEM	Sarno *et al.* 1997	***	***
Bacteriastrum elegans	**Na25A3**	GoN	Oct. 7, 2014	LM	Hustedt 1930	as Na29B1	as Na29B1
	Na27C1	GoN	Oct. 7, 2014	LM		NA	NA
	Na29B1#	GoN	Dec. 1, 2014	LM		***	***
	Na34C2	GoN	Jul. 28, 2015	LM		NA	NA
B. furcatum	PMF-BA4#	CA	May 10, 2011	LM, SEM, TEM	Bosak *et al.* 2015	KC914887	NA
B. hyalinum	**Na10B1#**	GoN	Mar. 19, 2014	LM, SEM	Bosak *et al.* 2015	***	***
	Na18A1	GoN	Jul. 1, 2014	LM		NA	NA
B. jadranum	PMF-BA1	MC	Oct. 17, 2010	LM, SEM, TEM	Bosak *et al.* 2015	KC914885	NA
	Na19C1#	GoN	Jul. 30, 2014	LM		***	***
	Na19C3	GoN	Jul. 30, 2014	LM		as Na19C1	as Na19C1
	Na25C4	GoN	Oct. 7, 2014	LM		NA	NA
	Na34A2	GoN	Jul. 28, 2015	LM		NA	NA
	Na35A1	GoN	Aug. 4, 2015	LM		NA	NA
B. mediterraneum	PMF-BA3	ZC	Apr. 28, 2011				
	Na1C4#	GoN	Nov. 26, 2013	LM	Bosak *et al.* 2015	as Na29B3	***
	Na4A4	GoN	Nov. 26, 2013	LM		as Na29B3	as Na1C4
	Na4C1	GoN	Nov. 26, 2013	LM		as Na29B3	as Na1C4
	Na17A1	GoN	Jul. 1, 2014	LM		NA	NA
	Na29B2	GoN	Dec. 1, 2014	LM		NA	NA
	Na29B3#	GoN	Dec. 1, 2014	NA		***	NA
B. parallelum	newLA1	GoN	Apr. 6, 2013	LM, SEM, TEM	Sarno *et al.* 1997	as newLA2	as newLA2
	newLA2#	GoN	Apr. 6, 2013	LM, SEM, TEM		***	***
Bacteriastrum sp.	**Na23A2**	GoN	Sep. 10, 2014	NA		NA	NA
Chaetoceros affinis V	SZN-B439#	GoN	Jul. 2004	LM, SEM	Kooistra *et al.* 2010	***	NA
	Na5B1	GoN	Nov. 26, 2013	LM		NA	NA

Strain name	Strain Code	Collection site [1]	Collection date	Images	Reference for morphological identification [2]	GenBank 28S rDNA	GenBank 18S rDNA
	Na5C4	GoN	Nov. 26, 2013	LM		NA	NA
	Na12B2	GoN	Mar. 19, 2014	LM		NA	NA
	Na17B3	GoN	Jul. 1, 2014	LM		NA	NA
	Na18A4	GoN	Jul. 1, 2014	LM		NA	NA
	Na31A2	GoN	Apr. 7, 2015	LM		NA	NA
	Na24A3	GoN	Oct. 2, 2014	LM		NA	NA
Chaetoceros affinis V	**Na24B4**	GoN	Oct. 2, 2014	LM		NA	NA
	Na27C2	GoN	Oct. 7, 2014	LM		NA	NA
C. affinis U	**Na2C2**	GoN	Nov. 26, 2013	LM		NA	NA
	Na25B3	GoN	Oct. 7, 2014	LM		NA	NA
	Na25B4	GoN	Oct. 7, 2014	LM		NA	NA
	Na28C1	GoN	Oct. 7, 2014	LM		NA	NA
C. cf. *affinis* V	**Na26B1#**	GoN	Oct. 7, 2014	LM, SEM		***	***
	Na26C2	GoN	Oct. 7, 2014	LM		as Na26C2	NA
C. anastomosans	**Na14A1#**	GoN	Mar. 19, 2014	LM, SEM	Shevchenko *et al.* 2006	***	NA
	Na14C2#	GoN	Mar. 19, 2014	NA		as Na14A1	***
	Na14C3	GoN	Mar. 19, 2014	LM		as Na14A1	as Na14C2
	Na14C4	GoN	Mar. 19, 2014	LM, SEM		as Na14A1	as Na14C2
	Na15C3	GoN	Apr. 24, 2014	NA		NA	NA
	Na28B2	GoN	Oct. 7, 2014	NA		as Na14A1	as Na14C2
	Na28B4	GoN	Oct. 7, 2014	LM		as Na14A1	NA
C. atlanticus	7C1#	SO	NA	LM, SEM, TEM	Kooistra *et al.* 2010	EF423455	***
C. brevis	**Na3C3**	GoN	Nov. 26, 2013	LM	Jensen & Moestrup 1998	as Na7B1	as Na7B1
	Na4A1	GoN	Nov. 26, 2013	LM		as Na7B1	as Na7B1
	Na7B1#	GoN	Jan. 1, 2014	LM		***	***
	Na34B2	GoN	Jul. 28, 2015	LM		as Na7B1	NA
	Na34B4	GoN	Jul. 28, 2015	LM		as Na7B1	NA
C. cf. *brevis*	**Na7C2#**	GoN	Jan. 17, 2014	LM, SEM		***	***
C. brevis/pseudobrevis	**Ch9B3#**	Con	Oct. 29, 2013	LM, SEM, TEM		***	***
C. cf. *pseudobrevis*	**Na22A1**	GoN	Sep. 10, 2014	NA		NA	NA

Strain name	Strain Code	Collection site [1]	Collection date	Images	Reference for morphological identification [2]	GenBank 28S rDNA	GenBank 18S rDNA
C. 'brush'	**Na26B4**	GoN	Oct. 7, 2014	LM	NA	NA	NA
	Na26C1	GoN	Oct. 7, 2014	LM		as Na28A1	as Na28A1
	Na28A1#	GoN	Oct. 7, 2014	LM, SEM, TEM		***	***
	Na36A2	GoN	Aug. 26, 2015	LM		NA	NA
C. calcitrans	PCC537#	NA	NA	NA	NA	NA	AY485449
	CAL	NA	NA	NA	NA	***	***
C. castracanei	MM24-C1#	SO	NA	LM, SEM	Kooistra *et al.* 2010	EF423453	NA
C. constrictus	**Ch5A3**	LC	Oct. 16, 2014	LM	Jensen & Moestrup 1998	NA	NA
	Ch5A4	LC	Oct. 16, 2014	LM, SEM, TEM		NA	NA
C. cf. *constrictus*	**Ch3C3**	LC	Oct. 16, 2013	LM, SEM, TEM		as Ch12C1	NA
	Ch12C1#	LC	Nov. 4, 2013	LM, SEM		***	***
	Ch13B1	LC	Nov. 5, 2013	LM		NA	NA
C. cf. *difficilis*	**Na2A3**	GoN	Nov. 26, 2013	LM	Hustedt 1930	NA	as Na18C4
	Na18A2	GoN	Jul. 1, 2014	LM		NA	as Na18C4
	Na18C4#	GoN	Jul. 1, 2014	LM		***	***
	Na24A2	GoN	Oct. 2, 2014	LM		NA	NA
	Na26A2	GoN	Oct. 7, 2014	LM		as Na18C4	as Na18C4
	Na26A4	GoN	Oct. 7, 2014	LM		as Na18C4	as Na18C4
	Na28B3	GoN	Oct. 7, 2014	LM		as Na18C4	as Na18C4
	Na34A4	GoN	Jul. 28, 2015	LM		NA	NA
C. cf. *holsaticus*	DH21#	GoN	Jun. 2004	LM	Jensen & Moestrup 1998	EF423433	NA
	Na13C1#	GoN	Mar. 19, 2014	LM	NA	as DH21	***
C. cf. *laevisporus*	Naos22#	GoP	NA	SEM	Kooistra *et al.* 2010	EF423437	NA
C. cf. *minimus*	MC755#	GoN	NA	NA	NA	***	NA
C. cf. *neogracile*	ArM0005#	NA	NA	NA	NA	NA	EU090014
C. cf. *pseudocrinitus*	**Na9A3**	GoN	Mar. 19, 2014	LM	Jensen & Moestrup 1998	as Na12A3	NA
	Na10A2	GoN	Mar. 19, 2014	LM		as Na12A3	NA
	Na11B1	GoN	Mar. 19, 2014	LM		as Na12A3	NA
	Na12A3#	GoN	Mar. 19, 2014	LM		***	***
	Na14A2	GoN	Mar. 19, 2014	LM, SEM		as Na12A3	NA

Strain name	Strain Code	Collection site [1]	Collection date	Images	Reference for morphological identification [2]	GenBank 28S rDNA	GenBank 18S rDNA
	Na31A1	GoN	Apr. 7, 2015	LM		NA	NA
	Na31B1	GoN	Apr. 7, 2015	LM		NA	NA
C. cf. *simplex*	CDP22#	GoN	NA	NA	NA	***	NA
C. cinctus	**Ch3A1**	LC	Oct. 16, 2013	LM	Cupp 1943	as Ch10B1	as Ch10B1
	Ch3C4	LC	Oct. 16, 2013	LM		as Ch10B1	as Ch10B1
	Ch6A2	LC	Oct. 16, 2013	LM		as Ch10B1	as Ch10B1
	Ch10B1#	LC	Oct. 29, 2013	LM		***	***
	Ch10B3	LC	Oct. 29, 2013	LM, SEM, TEM		as Ch10B1	as Ch10B1
	Ch10B4	LC	Oct. 29, 2013	LM		NA	NA
	Ch11C3	LC	Nov. 1, 2013	LM		as Ch10B1	as Ch10B1
C. circinalis	**Na15C2#**	GoN	Apr. 24, 2014	LM, SEM	Jensen & Moestrup 1998	***	***
C. compressus	clone6#	LI	Mar. 4, 2009	NA	Chamnansinp *et al.* 2015	KP300017	NA
C. compressus var *hirtisetus*	clone120#	LI	Mar. 4, 2009	LM, SEM, TEM		KP300024	NA
C. contortus	**Ch2B2**	LC	Oct. 16, 2013	LM		as Ch8A1	as Ch8A1
	Ch8A1#	Con	Oct. 29, 2013	LM, SEM, TEM		***	***
	Ch9A2	Con	Oct. 29, 2013	LM		as Ch8A1	as Ch8A1
	Ch11B3	LC	Nov. 1, 2013	LM		as Ch8A1	NA
	Ch12A4	LC	Nov. 4, 2013	LM		as Ch8A1	as Ch8A1
C. contortus var *ornatus*	clone18#	LI	Dec. 10, 2009	LM, SEM, TEM		KP300023	NA
C. contortus cf. var *contortus*	SZN-B402#	GoN	NA	LM, SEM		GU911462	NA
	Na31B2	GoN	Apr. 7, 2015	LM		as SZN-B402	NA
C. contortus cf. var *contortus*	DH22#	GoN	Jun. 2004	LM		EF423429	NA
	Na30A1#	GoN	Mar. 10, 2015	LM		NA	***
	Na30A2	GoN	Mar. 10, 2015	LM		NA	as Na30A1
	Na30B2	GoN	Mar. 10, 2015	LM		NA	NA
	Na31A4	GoN	Apr. 7, 2015	NA		NA	NA
C. convolutus	**Ch5C3#**	LC	Oct. 16, 2013	LM, SEM	Jensen & Moestrup 1998	***	***
	Ch5C4	LC	Oct. 16, 2013	LM, SEM		as Ch5C3	as Ch5C3
C. cf. *convolutus*	L7-B6#	NA	NA	NA		***	***
C. costatus	DH10#	GoN	May. 1, 2004	NA	Kooistra *et al.* 2010	EF423471	NA

Strain name	Strain Code	Collection site [1]	Collection date	Images	Reference for morphological identification [2]	GenBank 28S rDNA	GenBank 18S rDNA
	Na1A2	GoN	Nov. 26, 2013	LM		as DH10	NA
	Na1A3	GoN	Nov. 26, 2013	LM		as DH10	NA
	Na7A4	GoN	Jan. 17, 2014	LM		as DH10	as Na7C3
	Na7C3#	GoN	Jan. 17, 2014	LM		as DH10	***
	Na32B1	GoN	May. 22, 2015	LM		NA	NA
	Ro1B1	Roc	Aug. 11, 2014	LM		as DH10	as Na7C3
	Ro1C1	Roc	Aug. 11, 2014	LM		as DH10	NA
	Ro2A2	Roc	Aug. 11, 2014	LM		as DH10	NA
	Ro3B3	Roc	Aug. 11, 2014	LM		NA	NA
*C. curvisetus*1	DH8#	GoN	May. 1, 2004	NA	Kooistra *et al.* 2010	EF423477	NA
	Na1C1	GoN	Nov. 26, 2013	LM, SEM		as DH8	***
	Na20A4	GoN	Jul. 29, 2014	LM		as DH8	NA
*C. curvisetus*2	newBB2#	GoN	Apr. 6, 2013	LM		***	***
	Na3C4	GoN	Nov. 26, 2013	LM, SEM		as newBB2	NA
*C. curvisetus*3	**Na10B4#**	GoN	Mar. 19, 2014	LM, SEM		***	***
	Na10C1	GoN	Mar. 19, 2014	LM		as Na10B4	as 10B4
*C. curvisetus*4	**Ch5B1#**	LC	Oct. 16, 2013	LM		***	***
	Ch5B2	LC	Oct. 16, 2013	LM, SEM, TEM		as Ch5B1	as Ch5B1
	Ch12A2	LC	Nov. 4, 2013	LM		NA	as Ch5B1
C. cf. *curvisetus*	**Na19A2**	GoN	Jul. 30, 2014	LM		NA	NA
	Ro3B2	Roc	Aug. 11, 2014	LM		NA	NA
	Ro1B4	Roc	Aug. 11, 2014	LM		NA	NA
	Na32B4	GoN	May. 22, 2015	LM		NA	NA
	Na32C3	GoN	May. 22, 2015	LM		NA	NA
	Na33B2	GoN	Jul. 14, 2015	NA		NA	NA
	Na33C2	GoN	Jul. 14, 2015	LM		NA	NA
C. danicus	MC662#	GoN	NA	NA	Kooistra *et al.* 2010	EF423448	NA
	Na9B4#	GoN	Mar. 19, 2014	LM		***	***
C. dayaensis	MC107L#	DB	Oct. 12, 2012	LM, SEM, TEM	Li *et al.* 2015	KM401851	NA
C. cf. *dayaensis*	**Na11C3#**	GoN	Mar. 19, 2014	LM, SEM		***	***

Strain name	Strain Code	Collection site [1]	Collection date	Images	Reference for morphological identification [2]	GenBank 28S rDNA	GenBank 18S rDNA
C. debilis	**Ch1A1#**	LC	Oct. 16, 2013	LM	Kooistra *et al.* 2010	***	***
	Ch1B1	LC	Oct. 16, 2013	LM		as Ch1A1	as Ch1A1
	Ch2B3	LC	Oct. 16, 2013	LM		as Ch1A1	as Ch1A1
	Ch9A3	Con	Oct. 29, 2013	LM, SEM		as Ch1A1	NA
	Ch13A4	LC	Nov. 5, 2013	LM		as Ch1A1	as Ch1A1
	Ch13C3	LC	Nov. 5, 2013	LM		as Ch1A1	NA
	SEED2#	PO	NA	SEM, TEM		EF423483	NA
	CCMP164#	IOC	Jul. 3, 1979	NA		NA	KF899834
	MM24-A3#	SO	NA	NA		EF423485	***
	Ch48#	Japan	Oct. 8, 2008	NA		NA	AB847419
C. decipiens	DH2#	GoN	May. 1, 2004	SEM	Kooistra *et al.* 2010	EF423434	NA
	Na1A4	GoN	Nov. 26, 2013	LM		as DH2	as Na12B4
	Na2B4	GoN	Nov. 26, 2013	LM		NA	NA
	Na11B3	GoN	Mar. 19, 2014	LM, SEM		as DH2	as Na12B4
	Na12B4#	GoN	Mar. 19, 2014	LM		as DH2	***
	Na14B3	GoN	Mar. 19, 2014	LM		as DH2	as Na12B4
	Na18B4	GoN	Jul. 1, 2014	LM		NA	NA
	Na27B1	GoN	Oct. 7, 2014	LM		NA	NA
	Na28A2	GoN	Oct. 7, 2014	LM		as DH2	NA
	Na28A3	GoN	Oct. 7, 2014	LM		as DH2	NA
	Na33A3	GoN	Jul. 14, 2015	LM		NA	NA
	Na33B4	GoN	Jul. 14, 2015	NA		NA	NA
	Ro1B3	Roc	Aug. 11, 2014	LM		as DH2	NA
	Ro1C2	Roc	Aug. 11, 2014	LM		NA	NA
	Ro2A4	Roc	Aug. 11, 2014	LM		as DH2	NA
	Ro2C4	Roc	Aug. 11, 2014	LM		NA	NA
	RCC1997#	BS	Aug. 1, 2009	NA		JQ995413	JF794044
C. diadema	**Na9C1**	GoN	Mar. 19, 2014	LM	Jensen & Moestrup 1998	NA	NA
	Na12C1#	GoN	Mar. 19, 2014	LM		***	***
	Na13B1	GoN	Mar. 19, 2014	LM		as Na12C1	as Na12C1

Strain name	Strain Code	Collection site [1]	Collection date	Images	Reference for morphological identification [2]	GenBank 28S rDNA	GenBank 18S rDNA
	Na14B2	GoN	Mar. 19, 2014	LM		as Na12C1	as Na12C1
*C. diadema*1	**Ch1A4#**	LC	Oct. 16, 2013	LM	Kooistra *et al.* 2010	as Na12C1	***
	Ch1C1	LC	Oct. 16, 2013	LM		as Na12C1	as Ch1C1
	Ch4A1	LC	Oct. 16, 2013	LM, SEM, TEM		as Na12C1	as Ch1C1
*C. diadema*2	**Ch1A3#**	LC	Oct. 16, 2013	LM		***	***
	Ch5C1	LC	Oct. 16, 2014	LM, SEM, TEM		as Ch1A3	as Ch1A3
C. diadema-like	**Na33B3**	GoN	Jul. 14, 2015	LM		NA	NA
	Na16C4	GoN	Jul. 1, 2014	LM		NA	NA
	Ro3A4	Roc	Aug. 11, 2014	LM		NA	NA
	Ro3C3	Roc	Aug. 11, 2014	LM		NA	NA
*C. dichaeta*1	MM19-D3#	SO	NA	LM	Kooistra *et al.* 2010	EF423463	***
*C. dichaeta*2	MM19-A1#	SO	NA	LM		EF423465	***
C. dichatoensis sp. nov.	**Ch1B3**	LC	Oct. 16, 2013	LM, TEM	Chapter 5	as Ch9B4	as Ch9B4
	Ch4A4	LC	Oct. 16, 2013	LM, SEM, TEM		as Ch9B4	as Ch9B4
	Ch9B4#	Con	Oct. 29, 2013	LM, SEM, TEM		***	***
*C. didymus*1	**Ch2B4**	LC	Oct. 16, 2013	LM, SEM, TEM	Shevchenko *et al.* 2006	as Ch6A3	as Ch6A3
	Ch3B3	LC	Oct. 16, 2013	LM		as Ch6A3	as Ch6A3
	Ch6A3#	LC	Oct. 16, 2013	LM		***	***
*C. didymus*2	**Na20B4#**	GoN	Jul. 29, 2014	LM, SEM		***	***
C. diversus	DH28#	GoN	Jul. 2004	LM	Ruiz *et al.* 1993	EF423441	NA
	Na3C1	GoN	Nov. 26, 2013	LM, SEM		as DH28	as Na5B2
	Na5B2#	GoN	Nov. 26, 2013	LM		as DH28	***
	Na22B2	GoN	Sep. 10, 2014	LM		NA	NA
	Na15B4	GoN	Apr. 24, 2014	LM		as DH28	as Na5B2
	Na23B1	GoN	Sep. 10, 2014	LM		as DH28	NA
	Na15C1#	GoN	Apr. 24, 2014	LM		***	NA
C. eibenii	**Ch8C3**	Con	Oct. 29, 2013	LM	Jensen & Moestrup 1998	as Ro3A1	a Ro1B2
	Ro1B2#	Ros	Aug. 11, 2014	LM		as Ro3A1	***
	Ro3A1#	Ros	Aug. 11, 2014	LM		***	NA
	Ro3A2	Ros	Aug. 11, 2014	NA		NA	NA

Strain name	Strain Code	Collection site [1]	Collection date	Images	Reference for morphological identification [2]	GenBank 28S rDNA	GenBank 18S rDNA
	Ro3C2	Ros	Aug. 11, 2014	NA		NA	NA
C. fallax	CPH9#	NA	NA	LM	NA	KF219699	NA
C. gelidus	RCC1990#	BS	Jul. 1, 1905	NA	Chamnansinp *et al.* 2013	JQ995407	NA
	CNCIII51_13#	CAO	NA	NA		NA	HM581777
C. gracilis	ch.18#	NA	NA	NA	NA	NA	AY229897
C. holsaticus	CPH4#	China	Dec. 16, 2010	NA	NA	KP175042	NA
C. jonquieri	Tahiti#	NA	NA	NA	NA	NA	DQ830988
C. laciniosus	CPH21#	NA	NA	NA	NA	KF379751	NA
C. lauderi	DH12#	GoN	NA	LM	Kooistra *et al.* 2010	EF423431	NA
	Na1A1#	GoN	Nov. 26, 2013	LM		as DH12	***
	Na2A1	GoN	Nov. 26, 2013	LM		as DH12	as Na1A1
	Na13A4	GoN	Mar. 19, 2014	LM		as DH12	as Na1A1
	Na34C3	GoN	Jul. 28, 2015	LM		as DH12	NA
	Na34C4	GoN	Jul. 28, 2015	LM		as DH12	NA
	Na36A1	GoN	Aug. 26, 2015	LM		as DH12	NA
	Na36A3	GoN	Aug. 26, 2015	LM		as DH12	NA
C. lorenzianus	ITNADia51#	Japan	Mar. 16, 2007	NA	NA	NA	AB847414
C. cf. *lorenzianus*1	**Ch3B4#**	LC	Oct. 16, 2013	LM, SEM, TEM	Jensen & Moestrup 1998	***	***
	Ch4A3a	LC	Oct. 16, 2013	NA		as Ch3B4	NA
	Ch4C3	LC	Oct. 16, 2013	LM		as Ch3B4	as Ch3B4
	Ch5B4	LC	Oct. 16, 2013	LM		as Ch3B4	as Ch3B4
	Ch6B2	LC	Oct. 16, 2013	LM		as Ch3B4	as Ch3B4
	Ch11C1	LC	Nov. 1, 2013	LM		NA	as Ch3B4
C. cf. *lorenzianus*2	**Ch2C1**	LC	Oct. 16, 2013	LM		as Ch7B4	NA
	Ch3A3	LC	Oct. 16, 2013	LM		as Ch7B4	NA
	Ch3A4#	LC	Oct. 16, 2013	LM		as Ch7B4	***
	Ch3B2	LC	Oct. 16, 2013	LM		as Ch7B4	NA
	Ch4A3	LC	Oct. 16, 2013	LM		as Ch7B4	NA
	Ch4C4	LC	Oct. 16, 2013	LM		as Ch7B4	as Ch3A4
	Ch7B4#	LC	Oct. 16, 2013	LM, SEM, TEM		***	NA

Strain name	Strain Code	Collection site [1]	Collection date	Images	Reference for morphological identification [2]	GenBank 28S rDNA	GenBank 18S rDNA
	Ch11A1	LC	Oct. 31, 2013	LM		as Ch7B4	NA
	Ch11A2	LC	Oct. 31, 2013	LM		as Ch7B4	NA
	Ch13B4	LC	Nov. 5, 2013	LM		as Ch7B4	NA
C. cf. *lorenzianus*3	**Ch12A1#**	Con	Oct. 29, 2013	LM, SEM, TEM	Sunesen *et al.* 2008	***	***
C. muellerii	clone1134#	GM	Aug. 15, 2010	NA	NA	NA	JF790992
C. neogracile	CCMP163#	NA	NA	NA	NA	EF423469	NA
C. neogracile	AnM0002#	NA	NA	NA	NA	NA	EU090012
C. peruvianus	**Ch11B4**#	SA	Nov. 1, 2013	LM, SEM	Kooistra *et al.* 2010	***	***
	DH14#	GoN	NA	LM		EF423450	NA
	Na3A1	GoN	Nov. 26, 2013	LM		NA	NA
	Na4B4#	GoN	Nov. 26, 2013	LM		NA	***
	Na8A1	GoN	Feb. 6, 2014	LM		as DH14	as Na4B4
	Naos5#	GoP	NA	LM		EF423451	NA
C. cf. *peruvianus*	DS#	NA	NA	LM, SEM		***	NA
C. protuberans	SZN-B445#	GoN	NA	LM	Kooistra *et al.* 2010	EF423496	NA
	Ch8B1	Con	Oct. 29, 2013	LM		as SZN-B445	NA
	Ch8C2	Con	Oct. 29, 2013	LM, SEM, TEM		as SZN-B445	***
	Ch9C2	Con	Oct. 29, 2013	LM		as SZN-B445	as Ch8C2
	Na16B3	GoN	Jul. 1, 2014	LM		NA	NA
	Na30A4#	GoN	Mar. 10, 2015	LM		NA	as Ch8C2
	Ro1A2	Roc	Aug. 11, 2014	LM		as SZN-B445	NA
C. pseudocurvisetus	**Na13C4#**	GoN	Mar. 19, 2014	LM, SEM	Kooistra *et al.* 2010	***	***
	Na20C3	GoN	Jul. 29, 2014	LM		as Na13C4	as Na13C4
C. radicans	**Ch1B4**	LC	Oct. 16, 2013	LM, SEM, TEM	Sunesen *et al.* 2008	as Ch10A3	as Ch10A3
	Ch2A2	LC	Oct. 16, 2013	LM		as Ch10A3	as Ch10A3
	Ch3B1	LC	Oct. 16, 2013	LM		as Ch10A3	as Ch10A3
	Ch10A3#	LC	Oct. 29, 2013	LM, SEM		***	***
	Ch11A4	LC	Nov. 1, 2013	LM		as Ch10A3	as Ch10A3
	Ch10C1	LC	Oct. 29, 2013	LM		NA	NA
*C. radicans*2	CCMP197#	AO	NA	NA		AB430626	AB430592

Strain name	Strain Code	Collection site [1]	Collection date	Images	Reference for morphological identification [2]	GenBank 28S rDNA	GenBank 18S rDNA
C. rostratus	**Na1B3#**	GoN	Nov. 26, 2013	LM	Rines & Hargraves 1998	***	NA
	Na1C3#	GoN	Nov. 26, 2013	LM		as Na1B3	***
	Na6B3	GoN	Jan. 7, 2014	LM		as Na1B3	as Na1C3
	Na6B4	GoN	Jan. 7, 2014	LM		as Na1B3	as Na1C3
C. rotosporus	DY6#	DB	Dec. 16, 2010	LM	Li *et al.*2013	KC840041	NA
	Na18C1	GoN	Jul. 1, 2014	LM		NA	as Na22B1
	Na21A1	GoN	Sep. 2, 2014	NA		as Na22B1	NA
	Na21A2	GoN	Sep. 2, 2014	LM		as Na22B1	NA
	Na22A3	GoN	Sep. 10, 2014	LM		as Na22B1	as Na22B1
	Na22B1#	GoN	Sep. 10, 2014	LM		***	***
	Na22B3	GoN	Sep. 10, 2014	LM		as Na22B1	as Na22B1
	Na23A1	GoN	Sep. 10, 2014	LM		as Na22B1	NA
	Na23A3	GoN	Sep. 10, 2014	LM		as Na22B1	NA
	Na24B3	GoN	Oct. 2, 2014	LM, SEM		as Na22B1	NA
	Na27A4	GoN	Oct. 7, 2014	LM		as Na22B1	NA
C. seiracanthus	newEB3big#	GoN	Apr. 6, 2013	LM	Jensen & Moestrup 1998	***	***
*C. '*singlecells*'*	**Na17B2#**	GoN	Jul. 1, 2014	LM	NA	NA	***
	Na31A3	GoN	Apr. 7, 2015	LM		NA	NA
	Na31B4	GoN	Apr. 7, 2015	LM		NA	NA
C. socialis sensu stricto	No_1#	MI	Oct. 31, 2008	LM	Chamnansinp *et al.* 2013	KF219700	NA
	newEA3#	GoN	Apr. 6, 2013	LM		***	***
	Na2C4	GoN	Nov. 26, 2013	LM, SEM		as newEA3	NA
	Na4C4	GoN	Nov. 26, 2013	LM, SEM, TEM		as newEA3	NA
	Na12C2	GoN	Mar. 19, 2014	LM		as newEA3	NA
	Ro4A1	Roc	Aug. 11, 2014	LM		NA	NA
	Ro4A2	Roc	Aug. 11, 2014	NA		as newEA3	NA
	Na33A2	GoN	Jul. 14, 2015	NA		NA	NA
	Na33A4	GoN	Jul. 14, 2015	NA		NA	NA
	Na33B1	GoN	Jul. 14, 2015	LM		NA	NA
	Na33C3	GoN	Jul. 14, 2015	LM		NA	NA

Strain name	Strain Code	Collection site [1]	Collection date	Images	Reference for morphological identification [2]	GenBank 28S rDNA	GenBank 18S rDNA
C. sporotruncatus sp. nov	**Ch2A4#**	LC	Oct. 16, 2013	LM, SEM, TEM	Chapter 5	***	***
	Ch9C4	Con	Oct. 29, 2013	LM, SEM, TEM		as Ch2A4	as Ch2A4
C. tenuissimus	Chaet25.6#	NA	NA	LM, SEM, TEM	Kooistra *et al.* 2010	EF423470	NA
	Na26A1	GoN	Oct. 7, 2014	NA		as Chaet25.6	***
	Na14C1	GoN	Mar. 19, 2014	NA		as Chaet25.6	as Na26A1
C. teres	**Ch2C3**	LC	Oct. 16, 2013	LM, SEM, TEM	Shevchenko *et al.* 2006	as Ch6B1	as Ch6B1
	Ch5A1	LC	Oct. 16, 2013	LM		as Ch6B1	as Ch6B1
	Ch6A4	LC	Oct. 16, 2013	LM		as Ch6B1	as Ch6B1
	Ch6B1 #	LC	Oct. 16, 2013	LM		***	***
	Ch12B1	LC	Nov. 4, 2013	LM		as Ch6B1	as Ch6B1
	Ch13B2	LC	Nov. 5, 2013	LM		as Ch6B1	as Ch6B1
	Ro2B1	Roc	Aug. 11, 2014	LM		NA	NA
	Ro2B4	Roc	Aug. 11, 2014	NA		NA	NA
C. tortissimus	**Na25A2**	GoN	Oct. 7, 2014	LM	Shevchenko *et al.* 2006	as Na25B2	as Na25B2
	Na25B2#	GoN	Oct. 7, 2014	LM		***	***
	Na34B3	GoN	Jul. 28, 2015	LM		NA	NA
C. 'verylongsetae'	**Na13C2#**	GoN	Mar. 19, 2014	LM	NA	***	***
C. vixvisibilis	**Na16A3#**	GoN	Jul. 1, 2014	LM	Hernández-Becerril *et al.*2010	***	***
	Na16B2	GoN	Jul. 1, 2014	NA		NA	NA
	Na18C2	GoN	Jul. 1, 2014	LM, SEM		NA	NA
	Na25C3	GoN	Oct. 7, 2014	LM		as Na16A3	NA
	Na26B2	GoN	Oct. 7, 2014	LM		as Na16A3	as Na16A3
	Na28A4	GoN	Oct. 7, 2014	NA		as Na16A3	as Na16A3
	Na33A1	GoN	Jul. 14, 2015	LM		NA	NA
Chaetoceros sp.	Va7D2#	GoN	NA	LM, SEM	NA	***	NA
Chaetoceros sp.	StaVI#	NA	NA	LM	NA	NA	NA
Chaetoceros sp.	DH29#	GoN	NA	LM	NA	EF423439	NA
Chaetoceros sp.	ES9#	NA	NA	LM	NA	EF423432	NA
Chaetoceros sp.	clone7-G11#	GoF	NA	LM	NA	NA	FN690601
Chaetoceros sp.	**Na22A2**	GoN	Sep. 10, 2014	NA	NA	NA	NA

Strain name	Strain Code	Collection site [1]	Collection date	Images	Reference for morphological identification [2]	GenBank 28S rDNA	GenBank 18S rDNA
Chaetoceros sp.	**Na32C1**	GoN	May. 22, 2015	LM	NA	NA	NA
Chaetoceros sp.	**Na32A3**	GoN	May. 22, 2015	LM	NA	NA	NA
Chaetoceros sp.	**Na33C1**	GoN	Jul. 14, 2015	LM	NA	NA	NA
Chaetoceros sp.	**Na36B3**	GoN	Aug. 26, 2015	LM	NA	NA	NA
Chaetoceros sp.	**Na36B4**	GoN	Aug. 26, 2015	LM	NA	NA	NA
			Outgroups				
Ardissonea formosa	ECT3655	NA	NA	NA	NA	NA	HQ912653
Attheya longicornis	CCMP214	GoM	NA	NA	NA	NA	AY485450
Bellerochea malleus	NA	NA	NA	NA	NA	NA	AF525671
Biddulphiopsis titiana	ECT3697	NA	NA	NA	NA	NA	HQ912641
Calyptrella robusta	NA	NA	NA	LM		NA	AY485481
Cerataulina pelagica	ECT3845	NA	NA	LM		NA	HQ912669
Cymatosira belgica	NA	NA	NA	NA	NA	NA	X85387
Dactyliosolen blavyanus	Na14B4	GoN	Mar. 19, 2014	LM	NA	***	***
Dactyliosolen blavyanus	ECT3891	NA	NA	NA	NA	NA	KC309491
E. groenlandica	RCC1996	NA	NA	NA	NA	NA	JF794043
Eucampia sp.	Ch10C2	LC	Oct. 29, 2013	LM	NA	***	***
Eucampia groenlandica	RCC2037	NA	NA	NA	NA	JQ995430	NA
Eucampia zoodiacus	NA	NA	NA	NA	NA	GQ219682	NA
Eucampia zoodiacus	CCMP386	NA	NA	NA	NA	NA	EF585584
Helicotheca tamesis	NA	NA	NA	NA	NA	NA	X85385
Hemiaulus hauckii	DH20	GoN	NA	NA	NA	EF423428	NA
Hemiaulus sp	newAC3	GoN	NA	NA	NA	EF423428	***
Hydrosera sp.	CYTX025	NA	NA	NA	NA	NA	HQ912683
Isthmia enervis	CXCL	NA	NA	NA	NA	NA	HQ912684
Lithodesmium undulatum	CCMP1797	NA	NA	NA	NA	NA	HQ912559
Mastodiscus radiatus	ECT3822	NA	NA	NA	NA	NA	HQ912675
Odontella aurita	CCAP1054	NA	NA	NA	NA	NA	EU818943
Papiliocellulus simplex	CS431	NA	NA	NA	NA	NA	HQ912630
Porosira pseudodelicatula	CCMP1433	NA	NA	NA	NA	NA	AY485469

Strain name	Strain Code	Collection site [1]	Collection date	Images	Reference for morphological identification [2]	GenBank 28S rDNA	GenBank 18S rDNA
Terpsinoe musica	NHOP43	NA	NA	NA	NA	NA	HQ912682
Thalassiosira gessneri	ANO2-08	NA	NA	NA	NA	NA	DQ514864
Thalassiosira weissflogii	CCMP1010	NA	NA	NA	NA	NA	AY485445
Toxarium undulatum	WK44	NA	NA	NA	NA	NA	AF525668
Triceratium dubium	CCMP147	NA	NA	NA	NA	NA	HQ912572
Trigonium formosum	ECT3689	NA	NA	NA	NA	NA	HQ912648
Urosolenia eriensis	Y98NA8	NA	NA	NA	NA	NA	HQ912577

(1) Latitude and longitude of collection sites from my study are presented in Chapter 2. GoN refers to MareChiara site in the Gulf of Naples, Italy; LC- Las Cruces, Chile; Con- Concepción, Chile; Ros- Roscoff, Brittany, France; GoP- Gulf of Panama; GoM- Gulf of Maine; AO- Atlantic Ocean; BS- Beaufort Sea; CAO- central Arctic Ocean; GM- Gulf of Mexico; PO- north-western Pacific Ocean, 48.5°N 165°E; CA- central Adriatic, 43.1°N 16.3°E; MC- Maun Channel, northern Adriatic, 44.2°N, 14.5°E; ZC-Zadar Channel, northern Adriatic, 44.1°N, 15.1°E; SO- Atlantic Sector of the Southern Ocean, 02°E 49°S; DB- Daya Bay, South China Sea, 22.4°N, 114.4°E; MI- Mannai Island, Thailand, 12.4°N 101.4°E; LI- Loan Island, Phuket Province on Andaman Sea, Thailand, 08.0°N 98.5°E; IOC- Islas Orcadas Cruise, Station 6, 61.3°S 54.7°W;

(2) References for information on morphology and identification of Chaetocerotacean strains.

Results

In the present study, I characterised 270 strains belonging to the family Chaetocerotaceae isolated from the Italian, Chilean and French coasts (see Table 3.1 for strain codes, species names, morphological references, collection sites and dates, and GenBank numbers). Among these isolated strains, 251 belonged to *Chaetoceros* and 19 to *Bacteriastrum*.

a) Species description

LM and SEM illustrations of 44 morpho-species are portrayed in Figs 3.1−3.43.

***B.* cf. *furcatum* Shadbolt**

Fig. 3.1. a−i

The vegetative morphology of strain Na8A3 was examined. The gross morphology of this strain matched the description provided by Sarno *et al.* (1997) based on culture material obtained from the site. The ultrastructural detail was reproduced from Sarno *et al.* (1997, figs 19-34). Cells forming straight chains (Fig. 3.1. a). Cells cylindrical with several chloroplasts. Setae number variable (5−7), setae arise radially from the valve face margin, and fuse along a short distance with those of the adjacent cells and then diverge lying in a valvar plane (Fig. 3.1. c). Colonies hetero-polar, i.e., setae of the anterior terminal valve arise radially and form a regular curve in a counter-clockwise direction in valve view; whereas the setae on posterior terminal valves bend inside of the chain, and then toward the posterior end (Fig. 3.1. a−d). Setae circular in cross section and ornamented with irregularly distributed poroids (small perforations), small spines organised spirally (Fig. 3.1. h). Spines absent in the proximal part of the setae (Figs 3.1. f−g). Both the terminal and intercalary valves delicate and ornamented with numerous radially branched costae. Terminal valves with a central rimoportula (Figs 3.1. e−f). Girdle bands with several open intercalary bands and connecting bands; each band shows an undulate margin (Fig. 3.1. i).

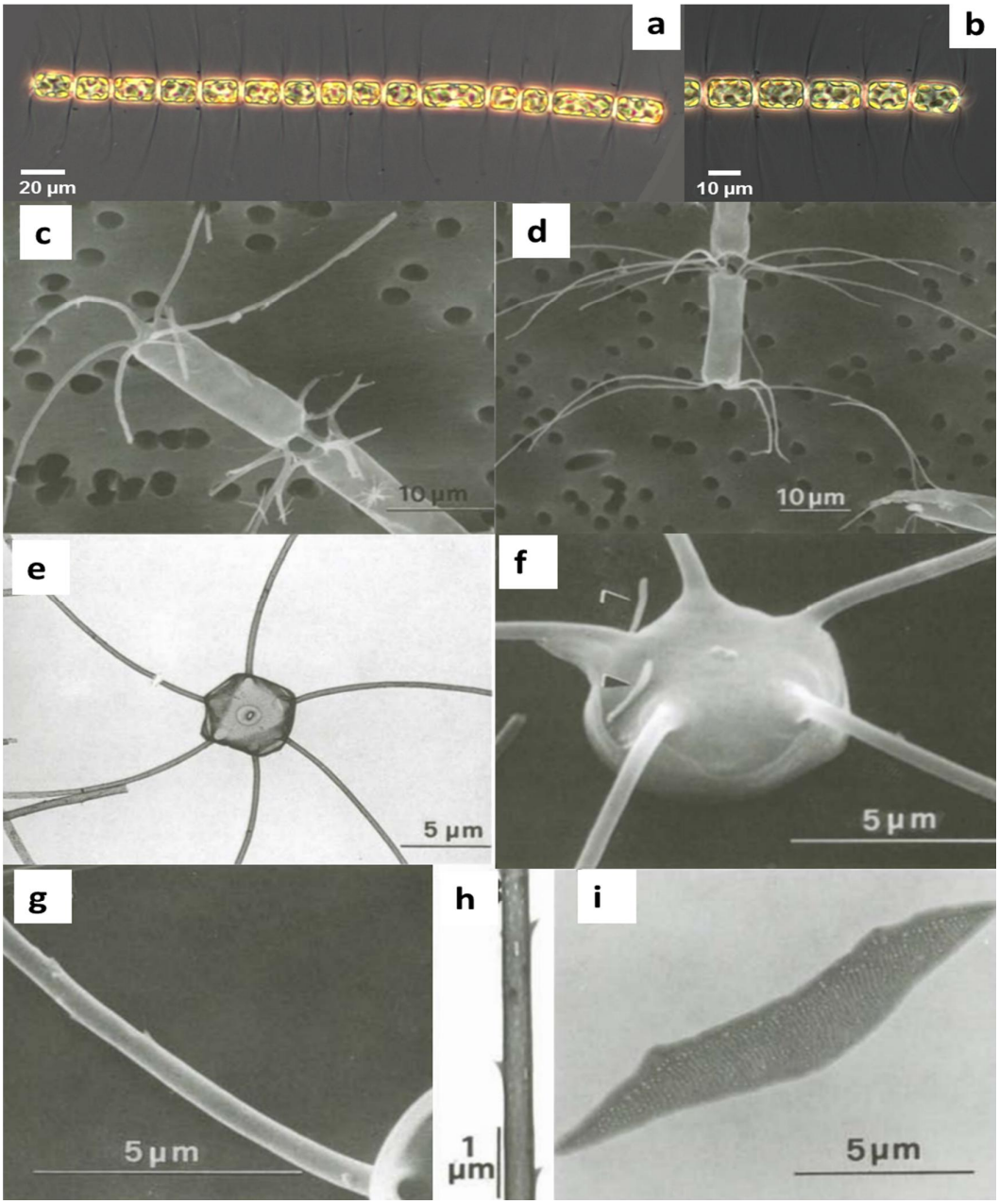


Fig. 3.1. *Bacteriastrum* cf. *furcatum*. Strain Na8A3 Figs (a–b). Fig. (c–i) are reproduced from Sarno *et al.* (1997). LM (a–b), SEM (c–d and f–g) and TEM (e and h–i). (a) A colony in girdle view, anterior pole on the left. (b) A part of a colony in girdle view with the posterior terminal cell. (c) Anterior terminal cell. (d) Posterior terminal cell. (e) Anterior terminal valve, with the rimoportula-like process and the central annulus. (f) Exterior detail of the posterior terminal valve. (g) Detail of the basal part of a seta. (h) Detail of a seta. (i) Intercalary band.

There was a taxonomic uncertainty regarding the actual *B. furcatum* because Bosak *et al.* (2015) published a 28S sequence of an Adriatic strain belonging to this species, which groups with *B. hyalinum* whereas the Neapolitan strain (Na8A3) is sister to *B. jadranum*. Such genetically markedly different but morphologically similar strains identified as *B. furcatum*

from the Tyrrhenian and Adriatic Sea (Bosak *et al.* 2015; present study) indicate the presence of cryptic species. However, in LM there are just very few characters to distinguish among the different *Bacteriastrum* species: the orientation of the terminal setae at both ends of a chain and the orientation of the intercalary setae. If these two taxa possess the states for these two characters, then they are considered to belong to the species, at least in LM. Observations in EM are needed to assess whether these strains differ in their ultrastructure.

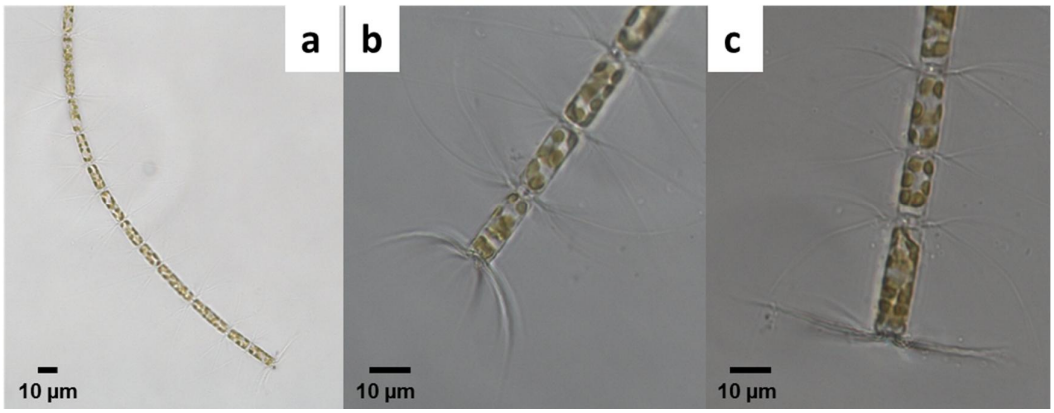


Fig. 3.2. *Bacteriastrum elegans*. Strain Na29B1 Figs (a−c). LM (a−c). (a) A colony in girdle view. (b) A part of a colony in girdle view with the posterior terminal cell. (c) Anterior terminal cell.

B. elegans **Pavillard**

Fig. 3.2. a−c

In this study, four strains of this species were isolated. The gross morphology of this strain matched the description provided by Hustedt (1930). The examined morphology belong to the strain Na29B1. Cells usually form straight or slightly curved chain (Fig. 3.2. a). Cylindrical cells with multiple chloroplasts. Setae number variable (5−7), arising radially from the valve face margin, and fuse along a short distance with those of the adjacent cells and then diverge in the pervalvar axis. Terminal setae thicker than the intercalary ones. The setae of the anterior terminal valve arise radially and form a regular curve in a counter-clockwise direction in valve view (Fig. 3.2. b). Also, the setae of the posterior terminal valves are thicker, but bend inside of the chain, and then toward the posterior end in an umbrella shape (Fig. 3.2. c). Ultrastructural details were not examined in this study.

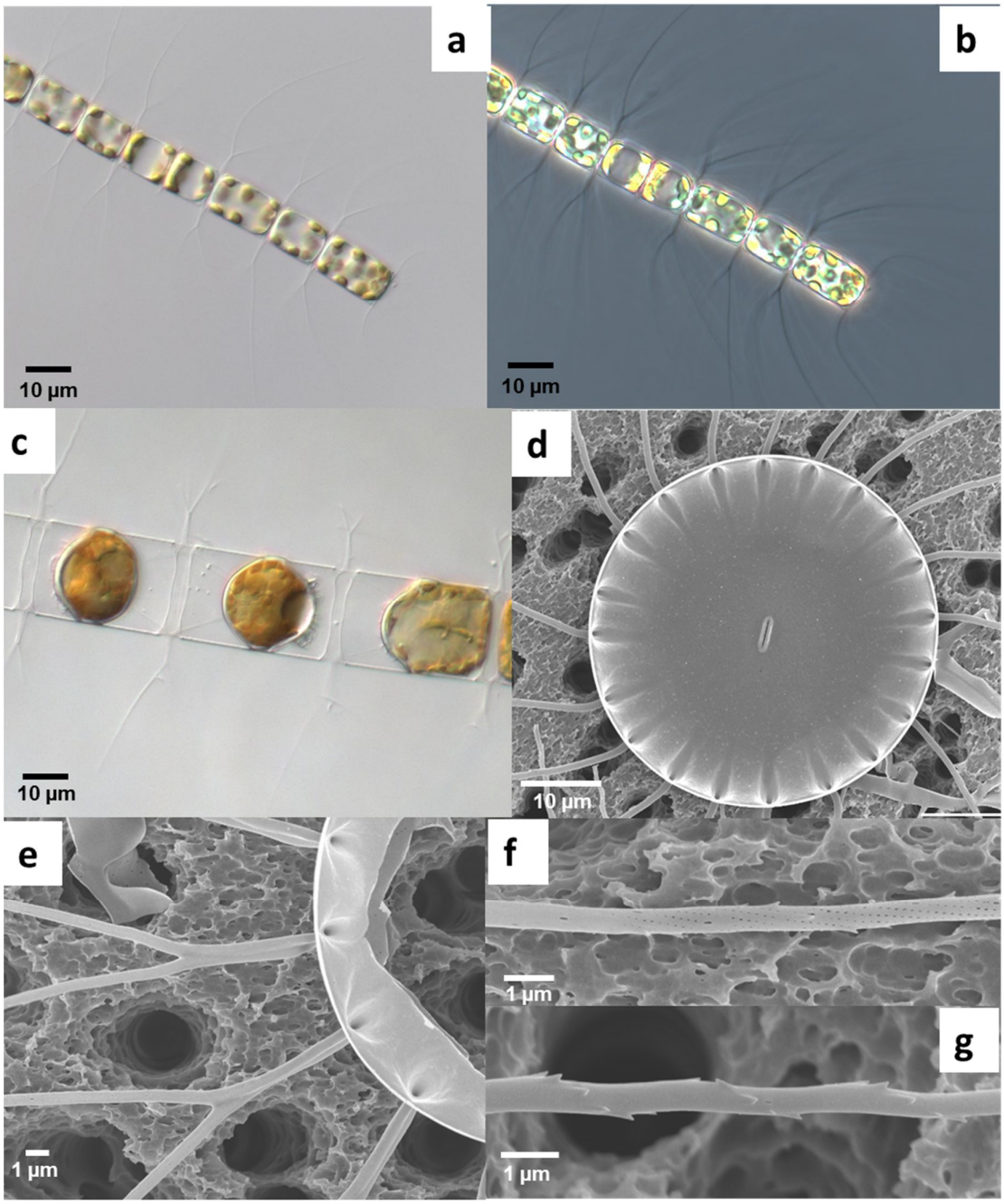


Fig. 3.3. *Bacteriastrum hyalinum*. Strain Na18A1 Figs (a−b) and Na10B1 Figs (c−g). LM (a−c), SEM (c−d and f−g) and TEM (e and h−i). (a) A colony in girdle view. (b) A part of a colony in girdle view with the posterior terminal cell. (c) Resting spore (d) Exterior detail of terminal valve, with the rimoportula-like process. (e) Detail of the basal part of a seta. (f−g) Details of setae.

***B. hyalinum* Lauder**

Fig. 3.3. a−g

In this study, two strains of this species were isolated. The gross morphology of this strain matched the description provided by Kooistra *et al.* (2010) and Bosak *et al.* (2015). The

morphology of the strains Na18A1 and Na10B1 were examined. Cells form short chains, with several chloroplasts (Fig. 3.3. a). Colonies isopolar, with similar anterior and posterior setae. Terminal setae umbrella-shaped (Fig. 3.3. b). Intercalary setae bifurcate after a short fusion for a distance and laid in the pervalvar plane (Fig. 3.3. e). Setae circular in cross section with spirally arranged spines and poroids, and regularly arranged large solitary pores (Figs 3.3. f–g). Valves circular in outline and contain several irregular small poroids. Terminal valve with a central elongated slit-shape, internally (Fig. 3.3. d). Endogenous round resting spores were seen in the culture (Fig. 3.3. c). Both valves ornamented with spines. A detailed description of this species is available in Bosak *et al.* (2015).

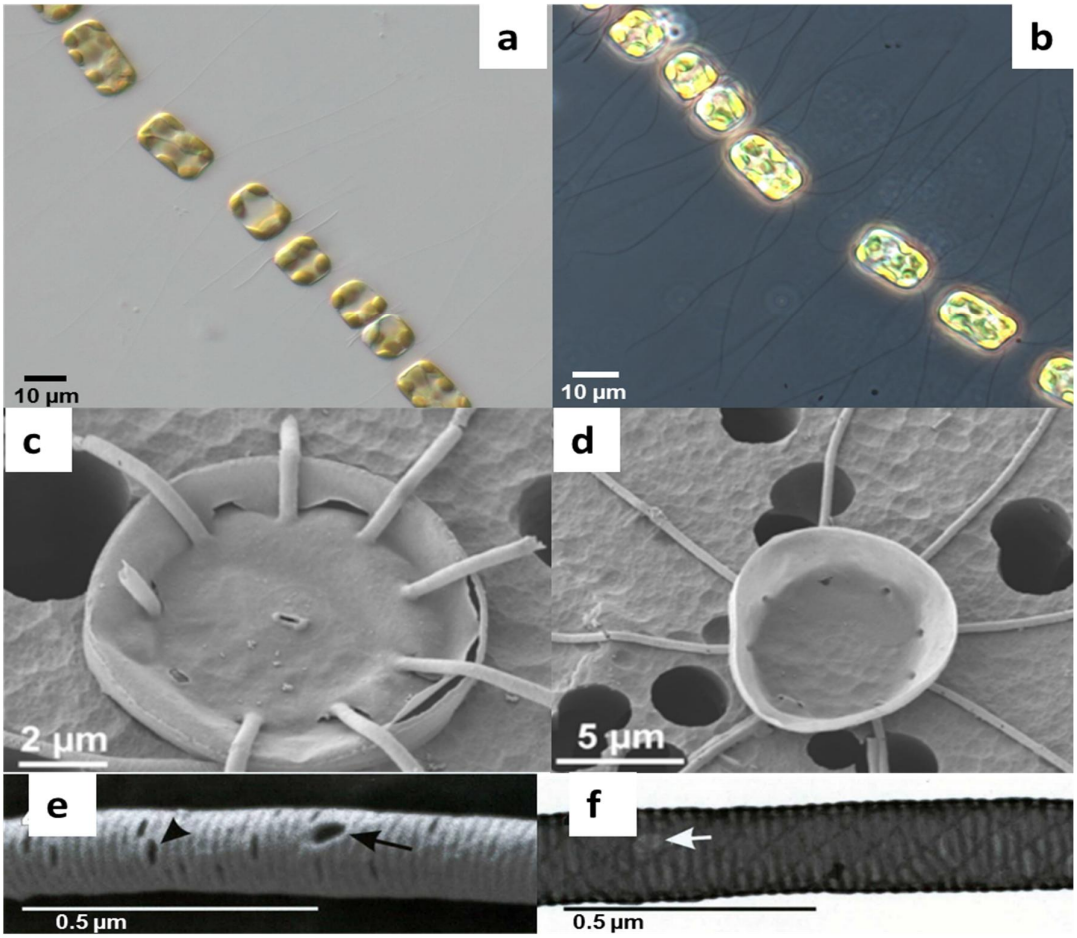


Fig. 3.4. *Bacteriastrum jadranum*. Strain Na19C3 Figs (a–b). Figs (c–d) and (e–f) are reproduced from Bosak *et al.* (2012) and (2015) respectively. LM (a–b), SEM (c–e) and TEM (f). (a) A colony in girdle view. (b) A part of a colony in girdle view with the posterior terminal cell. (c) Terminal valve with central slit-shaped rimoportula externally. (d) Single intercalary valve lacking central slit-shaped process. (e) Detail of a seta, presence of small pore (arrowhead) and elongated pore (arrow). (f) Detail of a seta (arrow mark large solitary pores).

B. jadranum **Godrijan, Marie & Pfannkuchen, emend. Bosak & Sarno**

Fig. 3.4. a–f

Five strains belonging to this morphotype were isolated. The gross morphology of this strain matched the description provided by Bosak *et al.* (2012, 2015). The partial 28S rDNA sequence generated in this study was identical to the sequence produced by Bosak *et al.* (2015). In this study, the morphology of the strain Na19C3 was examined and incorporated the ultrastructural details from Bosak *et al.* (2012, 2015). Cells forming long chains, aggregated in a mucilaginous membrane (Figs 3.4. a–b). Colonies isopolar, i.e. terminal setae on both sides similar. Setae originate from inside of valve margin (Fig. 3.4. c). Intercalary setae not fused. Setae smooth, perforated with both short and elongated poroids, and, without any spines (Figs 3.4. e–f). Terminal setae have similar morphology as of intercalary setae. Valve with a central annulus ornamented with radially branched costae. Terminal valve with a central process (Fig. 3.4. c). Intercalary cells without process (Fig. 3.4. d). Resting spores not reported.

B. mediterraneum **Pavillard**

Fig. 3.5. a–f

During this study, six strains corresponding to *B. mediterraneum* in LM were collected. The partial 28S rDNA sequence generated in this study was identical to the sequence produced by Bosak *et al.* (2015). Ultrastructure of this species was not examined but the morphology was identical to that reported in Bosak *et al.* (2015). Colonies straight with multiple chloroplasts (Fig. 3.5. a). Aperture between adjacent cells narrow. Colonies heteropolar, with anterior and posterior setae in a different orientation. Intercalary setae fuse for a distance and then bifurcate, diverging in the valvar plane (Fig. 3.5. d). Setae circular in cross-section, with spirally arranged spines and poroids, and equidistantly arranged large pores (Fig. 3.5. e). Valves circular in outline (Fig. 3.5. b). Terminal valves show a central slit-shape internally

and a small tube externally (Fig. 3.5. c). The cingulum is composed of several intercalary bands and one connecting band per valve. The bands possess one straight and one undulating margin (Fig. 3.5. f). Resting spores not reported. A detailed morphological description of the species is available in Bosak *et al.* (2015, see also Figs 52−67).

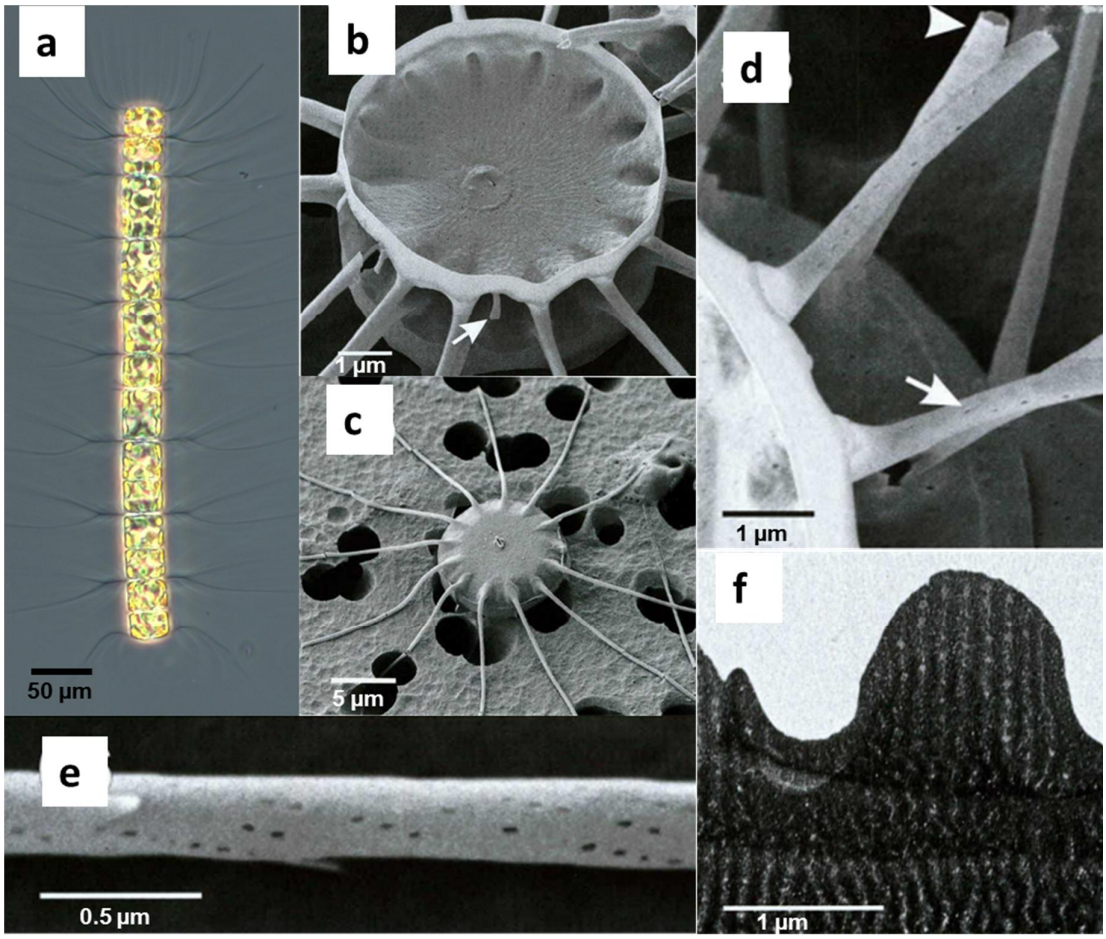


Fig. 3.5. *Bacteriastrum mediterraneum*. Strain Na1C4 Fig. (a). Figs (b−f) are reproduced from Bosak *et al.* (2015). LM (a), SEM (b−e) and TEM (f). (a) A colony in girdle view. (b) Sibling intercalary valves (arrow mark a shoehorn-shaped outgrowth projecting from a valve margin). (c) Terminal valve with central rimoportula. (d) Detail of setae basal part (arrow mark setae fusion, the surface of the valve and setae proximal part perforated by small pores). (e) Detail of a seta. (f) Detail of a single intercalary girdle band.

B. parallelum **Sarno, Zingone & D. Marino**

Fig. 3.6. a−g

This species was not collected in the present study and the morphology has been reproduced from Sarno *et al.* (1997, figs 3−18). Since the type locality is the GoN, the sequence obtained

from DNA isolated from a strain collected earlier at the MareChiara site was used. The description of this species is as follows: Solitary cells with multiple chloroplasts (Fig. 3.6. a). Species heterovalvate with 7−12 long setae (Figs 3.6. b−d). Setae of the anterior valve bend strongly towards the posterior valve, whereas the posterior valve setae make an acute bend and run parallel to the cell almost as anterior valve setae (Fig. 3.6. c).

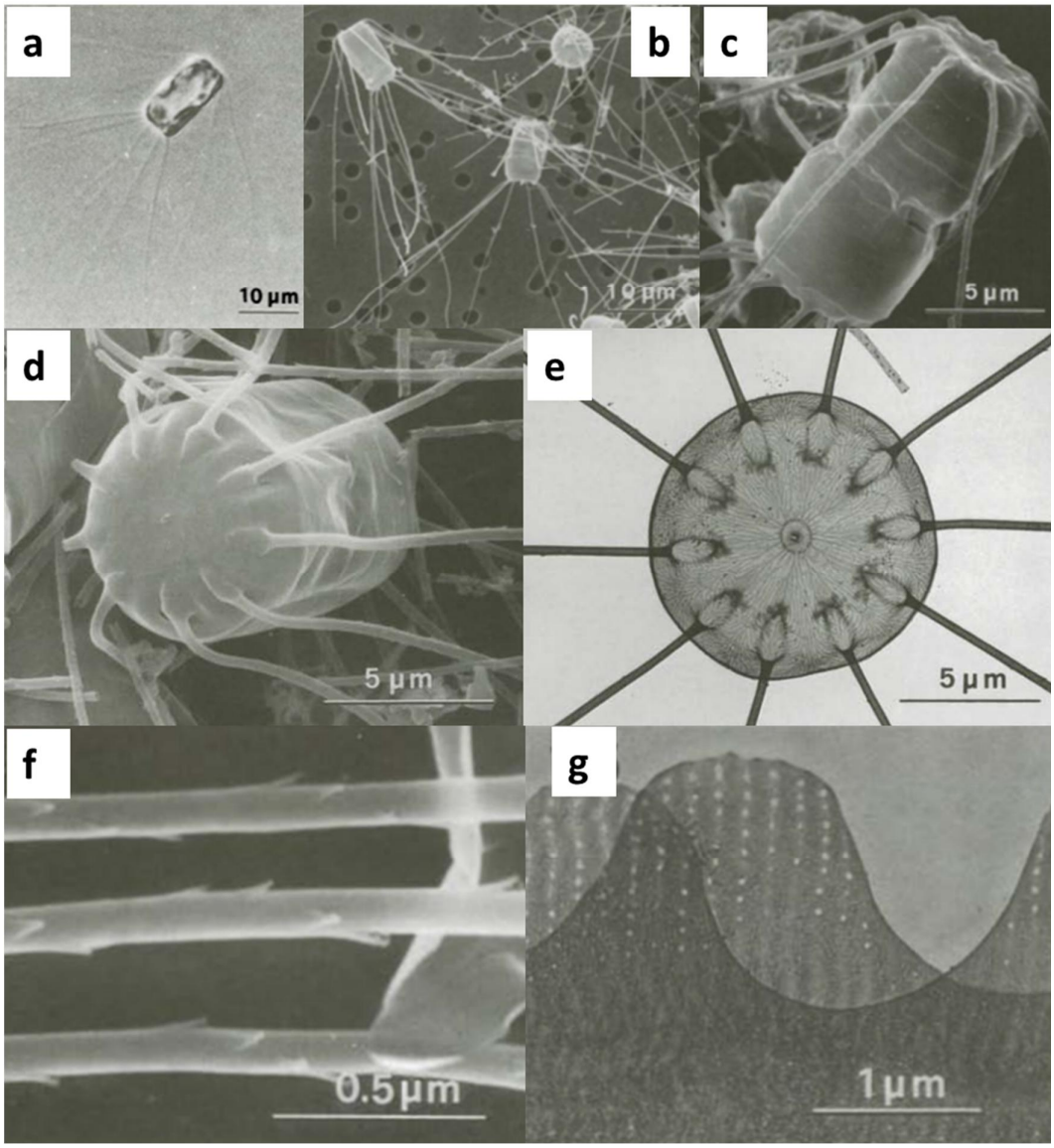


Fig. 3.6. *Bacteriastrum parallelum*. Figs (a−g) are reproduced from Sarno *et al.* (1997). LM (a), SEM (b−d and f) and TEM (e and g). (a) A colony in girdle view. (b) Sibling intercalary valves (arrow mark a shoehorn-shaped outgrowth projecting from a valve margin). (c) Detail of setae basal part (arrow mark setae fusion, the surface of the valve and setae proximal part perforated by small pores). (d) Terminal valve with central rimoportula. (e) Detail of setae. (f) Detail of a single intercalary girdle band.

Setae circular in cross section with spirally arranged spines and rows of poroids (Fig. 3.6. f). Valves circular in outline with costae radiating from the central annulus (Fig. 3.6. e). A small slit in present on the valve internally and externally it appears as a small tube (Fig. 3.6. e). The bands consist of regularly spaced rows of small pores parallel to the pervalvar axis (Fig. 3.6. g). Resting spores not known.

Chaetoceros affinis **Lauder**

Fig. 3.7. a−f

The strain described here was *C. affinis*, strain SZN-B439, for which morphological information and the partial 28S rDNA sequence were obtained by Kooistra *et al.* (2010). Cells forming straight chains with a narrow and lanceolate or slit-shaped aperture (Fig. 3.7. a). Each cell contains a single central plastid. Setae originate from the valve corners with a short basal part, crossing immediately and diverging sideways, but remaining in the apical plane. Intercalary setae circular in cross section, with spirally arranged shark-fin spines, and longitudinally organised poroids along the setae axis (Fig. 3.7. d). Terminal setae thicker than intercalary setae, curving smoothly and show dimorphism, i.e., the thin setae were identical to the intercalary setae structurally, and the thick setae polygonal in cross-section with longitudinal rows of poroids and spirally arranged spines (Figs 3.7. e−f). In very thick setae the poroids were not clearly visible. Valves perforated irregularly with very small poroids (Fig. 3.7. b). Rimoportula centrally located with a slit-shape on the interior side of the valve (Fig. 3.7. c) and with a tube-like structure externally (Fig. 3.7. c). Resting spores were not found in the culture material (Kooistra *et al.* 2010).

Several strains, identified as *C.* cf. *affinis* in LM, were collected during the study at LTER-MC, but it was not feasible to obtain the sequences for any of them. These strains could be sorted into three groups based on the shape of the terminal setae. The first group (including Na2C2, Na25B3, Na25B4, and Na28C1) had robust U-shaped pairs of terminal setae; a

second group possessed somewhat less robust V-shaped pairs of terminal setae (strains Na5B1, Na5C4, Na12B2, Na17B3, Na18A4, Na24A3, Na24B4, Na27C2, Na31A2 and the strain SZN-B439); and the third group possessed V-shaped terminal setae, which were however strongly diverging in the apical plane (strains Na26B1 and Na26C2). In the phylogenetic analysis, SZN-B439 was sister to *C. diversus* and Na26B1 was sister to *C.* cf. *dayaensis*. Only the examination of the type material, the examination of the ultrastructure (that could not be done in this thesis) and phylogenetic analyses will clarify the genetic relationships among these morphotypes. *Chaetoceros affinis* may be composed of a complex of several different species, each of which might fit one of the several varieties of this species reported in Guiry & Guiry (2016).

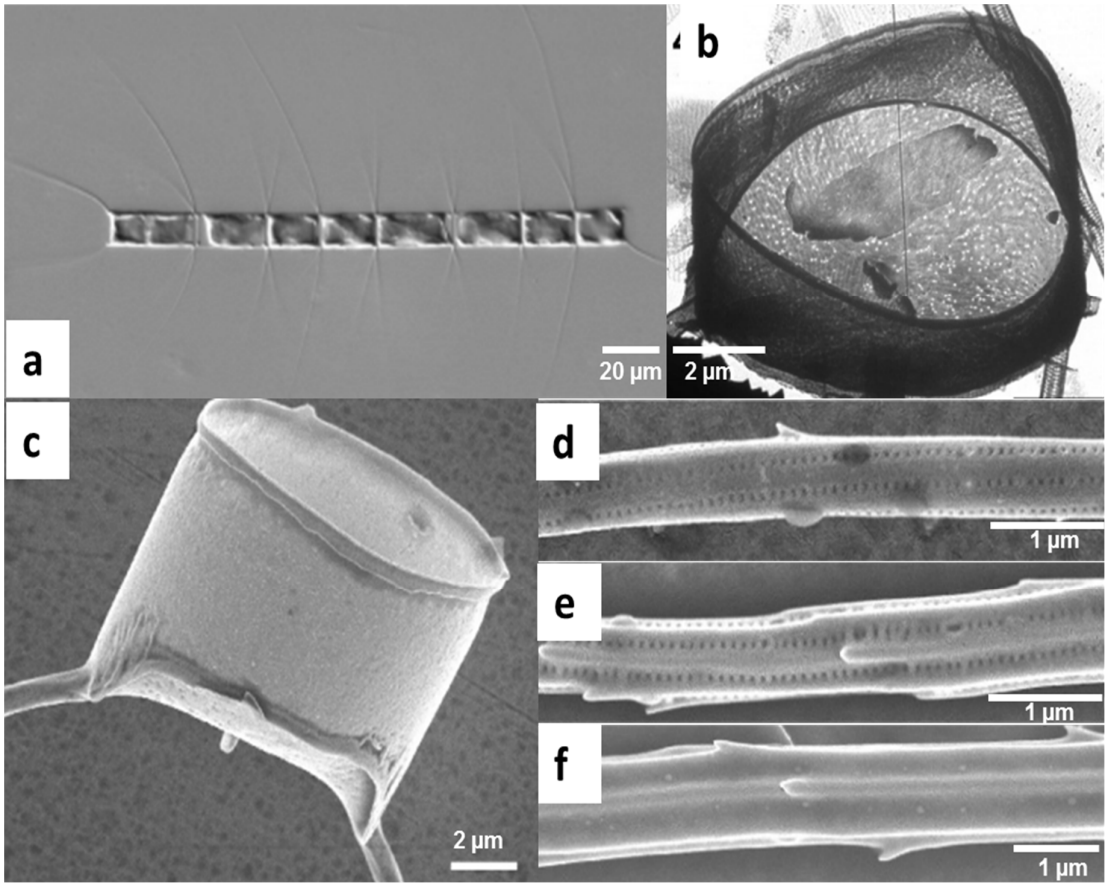


Fig. 3.7. *Chaetoceros* cf. *affinis*. Strain SZN-B439 Figs (a−g) are reproduced from Kooistra *et al.* (2010). LM (a), SEM (c−f) and TEM (b). (a) A colony in girdle view. (b) Terminal valve. (c) Terminal valve with a central tube-like rimoportula. (d) Detail of intercalary seta. (e−f) Details of terminal setae.

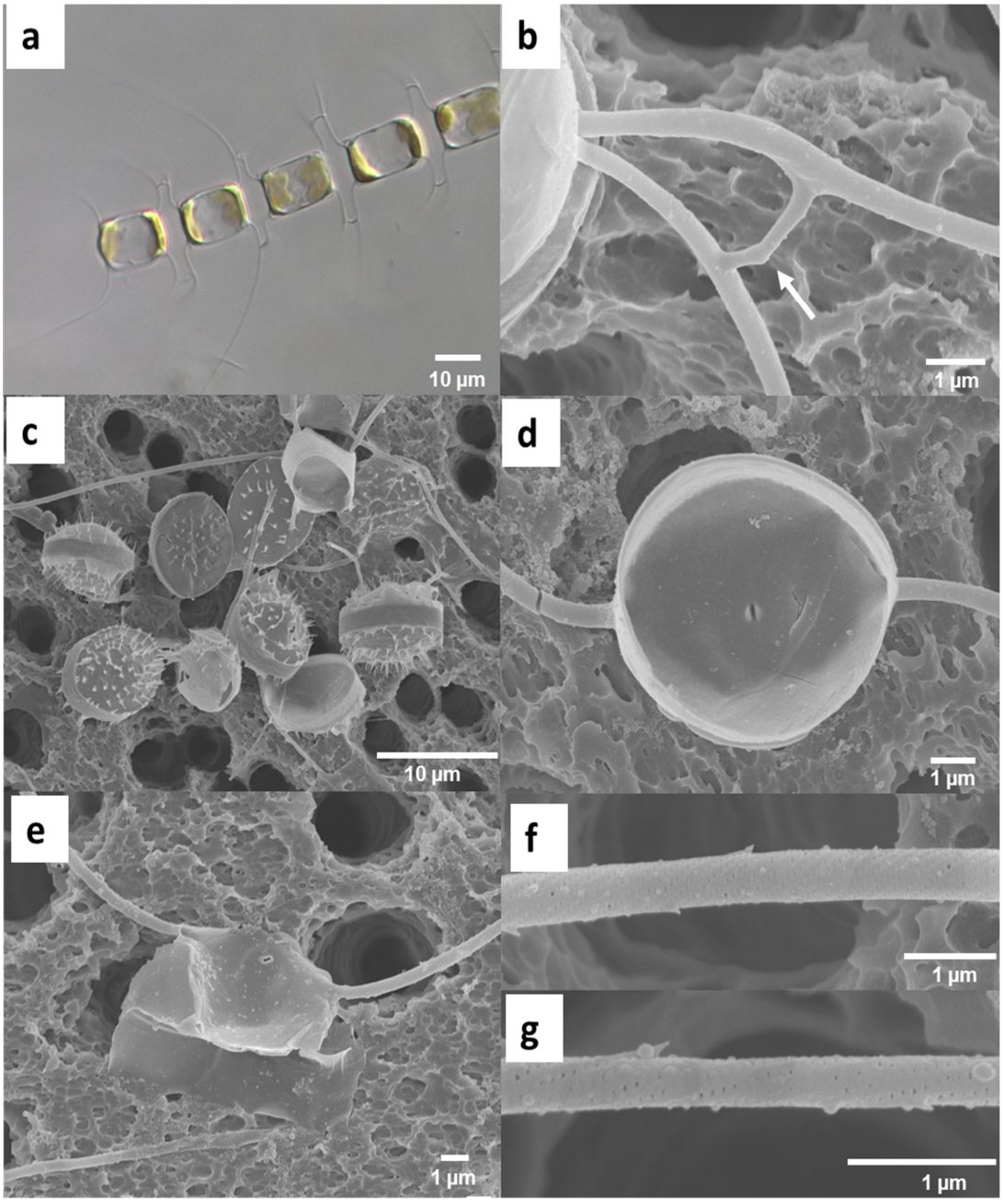


Fig. 3.8. *Chaetoceros anastomosans*. Strain Na14A1 Figs (a−g). LM (a), SEM (b-g). (a) A colony in girdle view. (b) Detail of the basal part of a seta (arrow indicate bridge between adjacent setae). (c) Resting spores. (d) Internal view of the terminal valve with a central slit-shaped rimoportula. (e) External view of the terminal valve with the central rimoportula. (f−g) Details of setae.

C. anastomosans **Grunow in Van Heurck**

Fig. 3.8. a−g

In this study, seven strains were isolated and morphologically identified as *C. anastomosans*.

The morphology of this strain matched the description provided by Shevchenko *et al.* (2006).

The illustrated morphology is that of strain Na14A1. Chains usually curved, with a wide octagonal aperture (Fig. 3.8. a). Each cell contains two plastids. Cells elliptical in valve view and rectangular in girdle view, with an unremarkable mantle. Setae long, with a long basal part. The setae do not cross over or fuse, but each one holds on to its counterpart of the adjacent cell by means of a silica bridge, and then diverges perpendicular to the chain axis (Fig. 3.8. b). Setae circular in cross section, with spirally arranged small shark fin spines and minute poroids (Figs 3.8. f–g). Valve elliptical, with costae radiating from the central annulus. The rimoportula is centrally located with a slit-shape on the interior side and small tube externally (Figs 3.8. d–e). Resting spores biconvex, with stout sharp spines on both valves (Fig. 3.8. c).

C. atlanticus **Cleve**

Fig. 3.9. a–e

This species was not collected in the present study, but a DNA sample collected by Kooistra *et al.* (2010) was used to obtain the 18S sequence. The morphological description belongs to the strain 7C1 from Kooistra *et al.* (2010). Colony straight and robust, with a wide hexagonal aperture (Fig. 3.9. a). Numerous small chloroplasts spread in the central body as well as in setae. Cells elliptical in valve view and rectangular in girdle view, with valves, slightly concave or flat, with a central undulation. Setae thick and long, with a short basal part (Fig. 3.9. b). The setae arise from valve corners, cross over chain axis, then diverge at an angle of 45° from chain axis and remain in an apical plane. Terminal setae parallel to the chain axis (Fig. 3.9. a). Solitary cells showed undulated setae (Fig. 3.9. c). Setae circular proximally in cross section, and later become polygonal distally to their junction bearing thick spines on the edges and rows of poroids on the sides (Fig. 3.9. e). Valve ornamented with irregular poroid patterns. All valves possess a central long tube like rimoportula externally (Fig. 3.9. b-d) and internally a simple hole. Resting spores not observed.

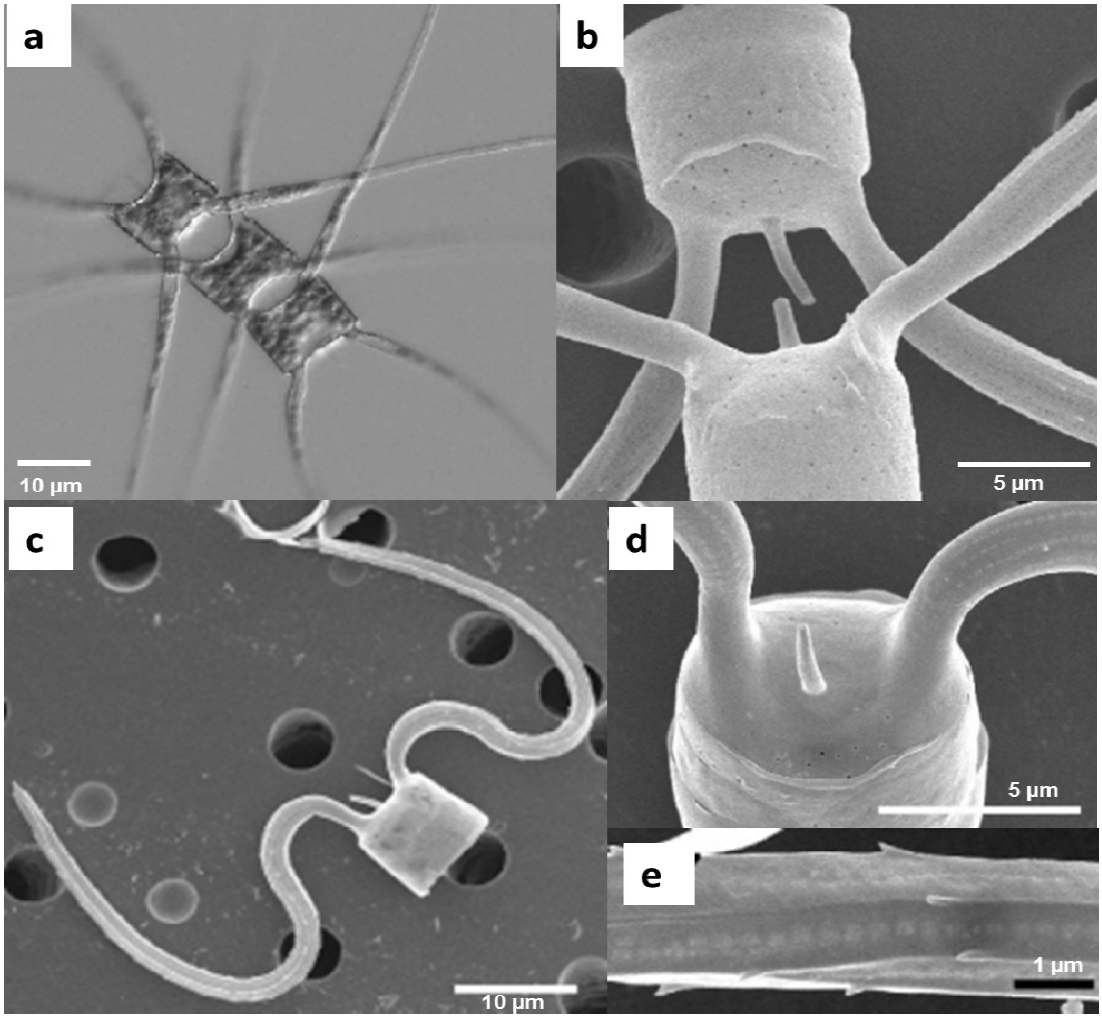


Fig. 3.9. *Chaetoceros atlanticus*. Strain 7C1 Figs (a−e) are reproduced from Kooistra *et al.* (2010). LM (a), SEM (b−e). (a) A colony in girdle view. (b) Sibling valves. (c) Terminal valves with a central tube like rimoportula; with diverging and curved setae. (d) Terminal valve with long external tube-shaped rimoportula. (e) Detail of a seta.

C. cf. *brevis* Schütt

Fig. 3.10. a−g

The examined morphology corresponds to strain Na7C2. Short chains, with a wide aperture (Fig. 3.10. a). One chloroplast per cell. Cells elliptical in valve view and rectangular in girdle view, with rounded apices. Mantle variable in height. Setae thin, delicate and arise from the inside of the margins with the long basal part, crossing immediately on chain axis. Setae contain small globules which appear as bright spots in LM. Intercalary setae circular in cross section, with spirally arranged spines and minute rows of poroids (Figs 3.10. f−g). Terminal

setae diverge at wide angle (Fig. 3.10. a). Valve elliptical to subcircular in outline, ornamented with branched costae. Terminal valve possesses an eccentric elongated rimoportula externally (Fig. 3.10. e), and large slit on the inside (Fig. 3.10. d). Resting spores biconvex, ornamented with several spines and have a collar on the secondary valve. This strain was designate as *C*. cf. *brevis*, because three different morphotypes belonging to *C. brevis* were obtained, which are genetically distinct. This strain (Na7C2) was similar to the description of *C. brevis* in Jensen & Moestrup (1998) and Sunesen *et al.* (2008), but its setae have a different orientation. Six more strains were isolated, among which five (Na3C3, Na4A1, Na7B1, Na34B2, Na34B4) match the description of *C. brevis* provided by Jensen & Moestrup (1998), and the latter strain (Ch9B3) differed from *C. brevis* by lacking the globules in the setae. The latter strains were labeled as *C. brevis/C. pseudobrevis.* Phylogenetic analyses showed that the three morphologically defined groups were genetically distinct but closely related. Further ultrastructural analyses are required to define differences among the three genotypes.

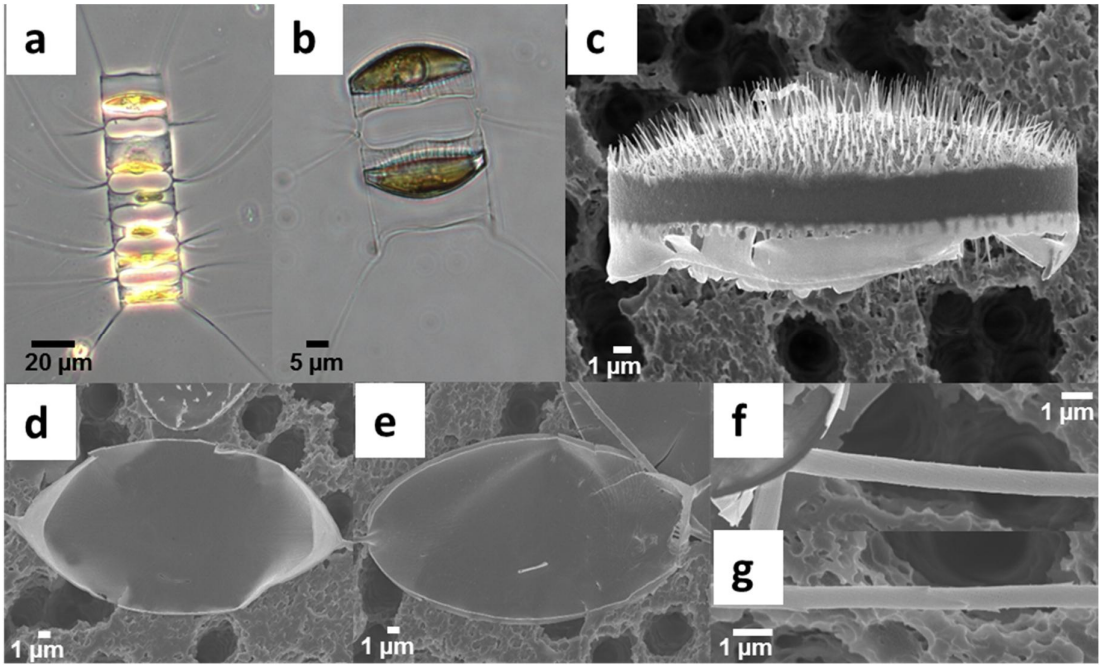


Fig. 3.10. *Chaetoceros* cf. *brevis*. Strain Na7C2 Figs (a–g). LM (a–b) and SEM (c–g). (a) A colony in girdle view. (b–c) Resting spores. (d) Terminal valves with a central slit-shaped rimoportula. (e) Terminal valve with short elongated external tube-shaped rimoportula, note the eccentric rimoportula. (f–g) Details of setae.

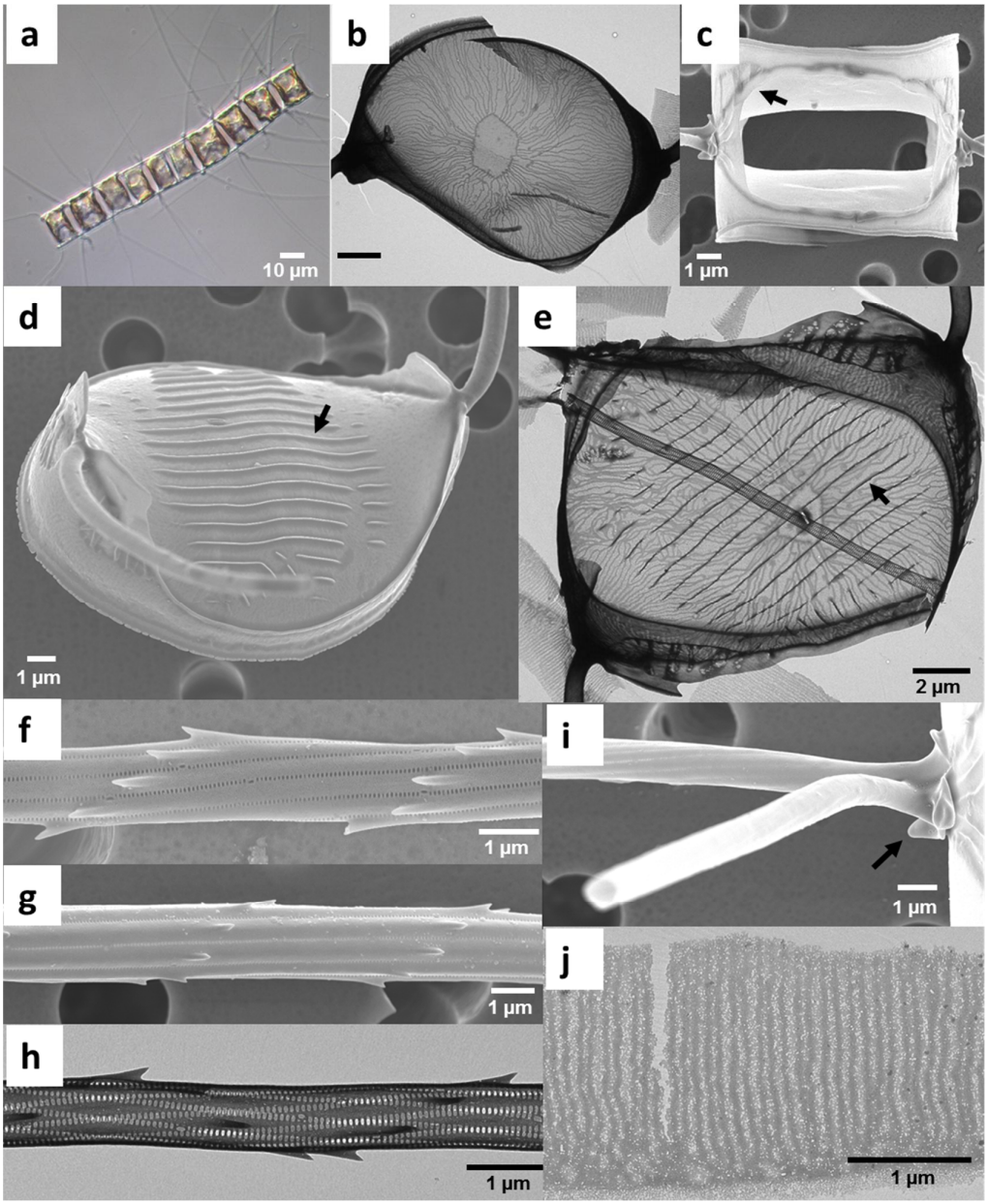


Fig. 3.11. *Chaetoceros* 'brush'. Strain Na28A1 Figs (a–j). LM (a), SEM (c–d, f–g and i) and TEM (b, e, h, and j). (a) A colony in girdle view. (b) Intercalary valve. (c) Aperture (arrow mark hyaline rim). (d–e) Terminal valves with the central slit-shaped rimoportula (arrow mark series of distinct parallel siliceous ridges) (f–h) Details of setae. (i) Petal-like collar at the base of the intercalary setae (arrow). (j) Girdle band.

C. 'brush'

Fig. 3.11. a–j

The examined and described morphology correspond to strain Na28A1, which differs from that of any described species and probably belongs to a species new to science. Cells usually in straight or curved chains (Fig. 3.11. a). In girdle view, cells appear elongated with drawn-

out corners, shallow mantle. Aperture narrow, elliptical to rectangular in shape. Two plate-like chloroplasts. Intercalary setae long and robust, arising from valve apices, protruding in all directions and fuse immediately with a single crossover point at the chain margin. Terminal setae long, straight and diverge slightly from each other extending parallel to the chain axis in a U shape (Fig. 3.11. a), sometimes crossing each other in distal part. Setae nonagonal or circular in cross-section. Setae pore narrow and run parallel to the shark fin-shaped spines, with large pores at irregular intervals (Figs 3.11. f−h).

Setae poroids small with a density of 126−130 poroids in 10 µm (*n*=10). Intercalary setae show a petal-like collar at their base (Fig. 3.11. i) along with a hyaline rim (Fig. 3.11. c). Valves appear broadly elliptical to circular, with dichotomous branching costae radiating from a central annulus (Fig. 3.11. b). Annulus shape variable from elliptical to elongate and sometimes irregular. Terminal valves with a slightly elevated central area. They show a series of distinct parallel siliceous ridges on the outer side of the terminal valve (Fig. 3.11. d−e). The species name 'brush' is provisional and refers to the morphology of the setae in LM that looks like a brush.

C. castracanei Karsten

Fig. 3.12. a−d

This species was not collected in the present study and the 28S sequence was from Kooistra *et al.* (2010). The morphological description of strain MM24C1 was obtained from Kooistra *et al.* (2010). Cells usually forming straight chains, apertures narrow (Fig. 3.12. a). Several small chloroplasts, present in the central body as well as in the setae. The setae arise from valve corners, cross over chain axis, and then diverge. Setae square in cross section, ornamented with thick spines and grids of pores (Figs 3.12. c−d). Valves ornamented with several rows of poroids and possess central rimoportula with a small external projection (Fig. 3.12. b). Resting spores not observed.

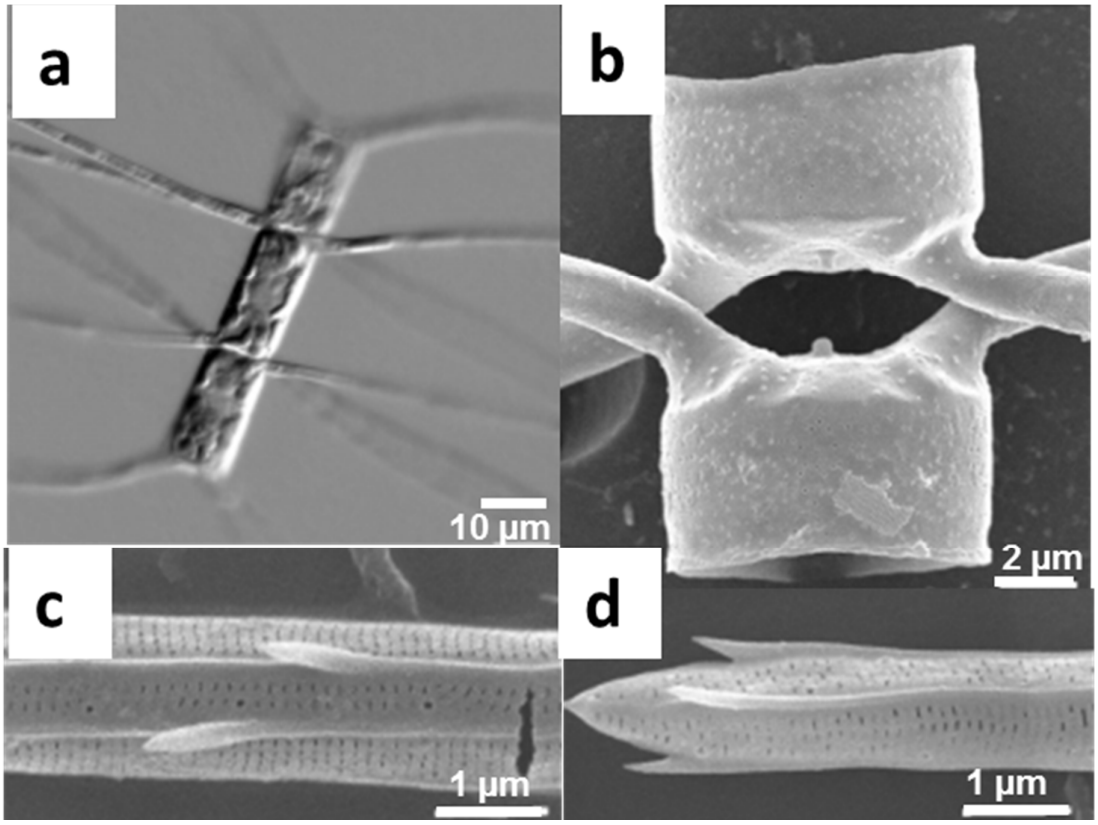


Fig. 3.12. *Chaetoceros castracanei*. Strain MM24C1 Figs (a−d) are reproduced from Kooistra *et al.* (2010). LM (a) and SEM (b−d). (a) A colony in girdle view. (b) Sibling valves. (c) Detail of the middle part of a seta. (d) Tip of a seta.

C. cf. *constrictus* Gran

Fig. 3.13. a−i

The examined morphology belongs to strain Ch12C1. Cells form chains with a narrow lanceolate aperture (Figs 3.13. a−b). Two chloroplasts visible per cell. Cells rectangular in girdle view. Mantle variable in height. Setae arise on the corners of the valve without a basal part, crossing immediately on chain axis (Fig. 3.13. b). Intercalary setae at the base of the chains bend towards one end of the colony (Fig. 3.13. b). Intercalary setae polygonal in cross section, with spirally arranged shark fin spine, and longitudinally arranged pores along chain axis, with regularly organised large pores (Fig. 3.13. h). Terminal setae parallel to the chain axis giving a V-shaped appearance, and ornamented with spirally arranged shark-fin spines, but longitudinally arranged slits (Fig. 3.13. i). Valves elliptical to sub-circular in outline, ornamented with branched costae. A small hyaline rim is present on the marginal ridge (Fig.

3.13. d). Terminal valve possesses several spines on valve face, along with a central tubular process externally (Figs 3.13. f–g), and internally a small slit (Fig. 3.13. e). Resting spores comprised of unequally convex valves (Fig. 3.13. c). Primary valve possesses long stout spines that are branched at the tips. The spines at the corner of the valve are conical and present branched pattern. The secondary valve had short stout spines but without bifurcation. The mantle of resting spore in the primary valve possesses small silica ornamentations.

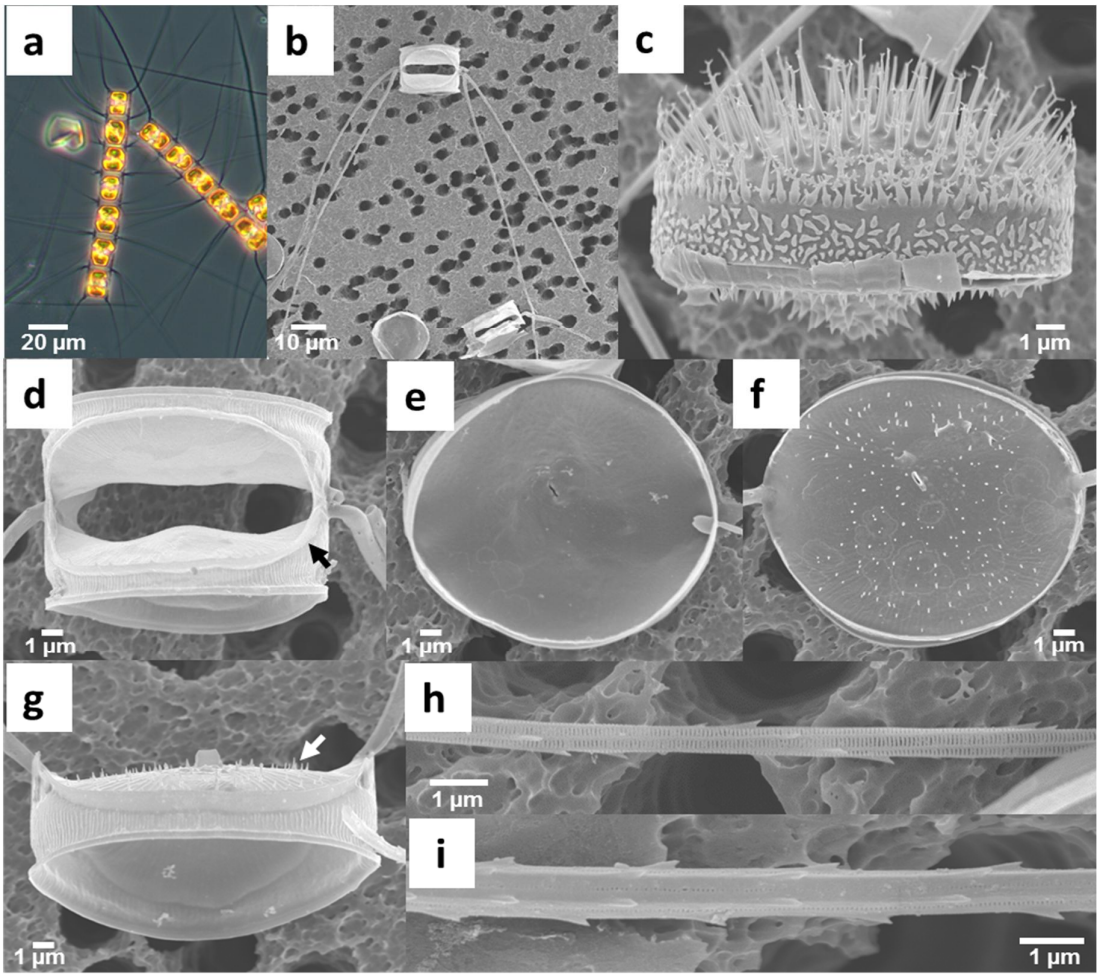


Fig. 3.13. *Chaetoceros* cf. *constrictus*. Strain Ch12C1 Figs (a–i). LM (a) and SEM (b–i). (a) A colony in girdle view. (b) Intercalary valve with setae bends towards one end of the chain. (c) Resting spore. (d) Two sibling valves with the arrow aperture (arrow mark hyaline rim). (e) Internal view of a terminal valve with the central slit-shaped rimoportula. (f–g) External view of terminal valves with the flattened tubular process of the rimoportula (arrow marks capilli on valve surface). (h) Detail of intercalary seta. (i) Detail of terminal seta.

In the present study, five strains representing similar morphologies were isolated. During the genetic characterisation, two strains (Ch5A3 and Ch5A4) failed repeatedly with all possible

primer combinations, whereas the other three strains (Ch3C3, Ch12C1, and Ch13B1) amplified without any difficulties. Therefore, these two groups of strains might belong to two distinct species. The ultrastructure of strain Ch12C1 matched with that reported by Jensen and Moestrup (1998) in terms of setae morphology and valve ornamentation, but the authors did not describe the resting spore. The ultrastructural details of the examined strain did however not match with that described in Shevchenko *et al.* (2006). Strain Ch12C1 showed setae polygonal in cross-section with spirally arranged shark-fin spines and longitudinally organised slits along the setae axis, whereas those illustrated in Shevchenko *et al.* (2006; figs 44-46) are rounded in cross-section. Also, the terminal valve does not show any spines in their study; hence, we indicated the examined strain as *C.* cf. *constrictus*.

C. cf. *dayaensis*

Fig. 3.14. a−f

One strain (Na11C3) was isolated during the study, whose gross morphology was similar to the description of *C. dayaensis* by Li *et al.* (2015). Chains are slightly twisted, with apertures of variable size (Fig. 3.14. a). Each cell contains one chloroplast. Intercalary setae long and emerge on the valve corners and cross just outside the valve margin, and are directed in all directions. Terminal setae straight, almost parallel to the axis of the colony, and generally diverging from the apical plane. Setae circular in cross-section and possess numerous minute spirally arranged poroids and small spines (Figs 3.14. e−f). Valves elliptical in outline and ornamented with dichotomously branched costae (Fig. 3.14. b). Terminal valve possesses a central slit-shaped rimoportula (Fig. 3.14. c) and a short external flattened tube (Fig. 3.14. d). Resting spores were not seen in the culture material. The examined strain differed from *C. dayaensis*, based on the ultrastructure of the setae. The setae of strain Na11C3 were circular in cross-section, while those of *C. dayaensis* had four- or six-sided setae (Li *et al.*

2015). The 28S rDNA sequence of the isolated strain differs from that of *C. dayaensis* and clusters in the clade that includes, among others, *C. anastomosans* and *C. vixvisibilis*.

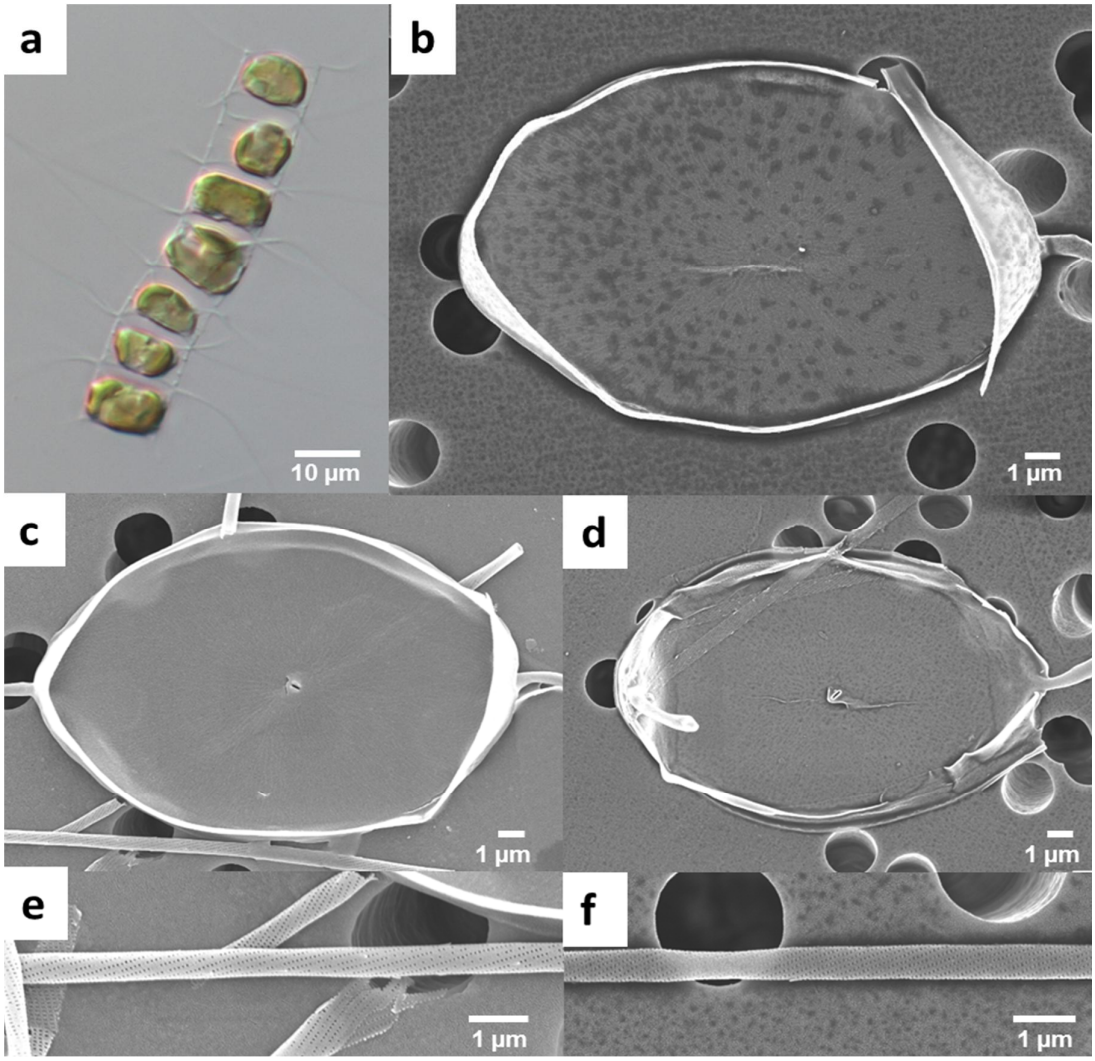


Fig. 3.14. *Chaetoceros* cf. *dayaensis*. Strain Na11C3 Figs (a−f). LM (a) and SEM (b−f). (a) A colony in girdle view. (b) Intercalary valve. (c) Internal view of a terminal valve with the central slit-shaped rimoportula. (d) External view of a terminal valve with the short flattened tube. (e−f) Details of setae.

***C. dayaensis* Li and Zhu**

Fig. 3.15. a−j

This strain was not collected in this study. The described morphology corresponds to the strains MC107L and MC107S from Li *et al.* (2015). Cells usually in straight or twisted chains, showing wide apertures with central inflation (Fig. 3.15. a). Each cell contains one chloroplast. Mantle low or equidimensional with the girdle, a slight constriction is visible

near the suture (Figs 3.15. d−e). Intercalary setae originate from the valve apices with the short basal part and cross just over the chain margin and are oriented in all directions. Terminal setae straight, almost parallel to the axis of the colony and generally diverge from the apical plane, giving a U-shaped appearance. Setae four or six-sided with rows of spines and elongated pores (Figs 3.15. h−i). Valve elliptical to semi-circular in outline, ornamented with radially branched costae. The terminal valve possesses a central slit-shaped rimoportula (Fig. 3.15. f) with a short external tube (Fig. 3.15. f). The maturation of resting spores is characterized by two rotations in the mother cell, similarly to what observed in *C. rotosporus*. Spores ornamented with small cone-shaped spines (Figs 3.15. b−c). Girdle bands ornamented with transverse parallel costae (Fig. 3.15. j).

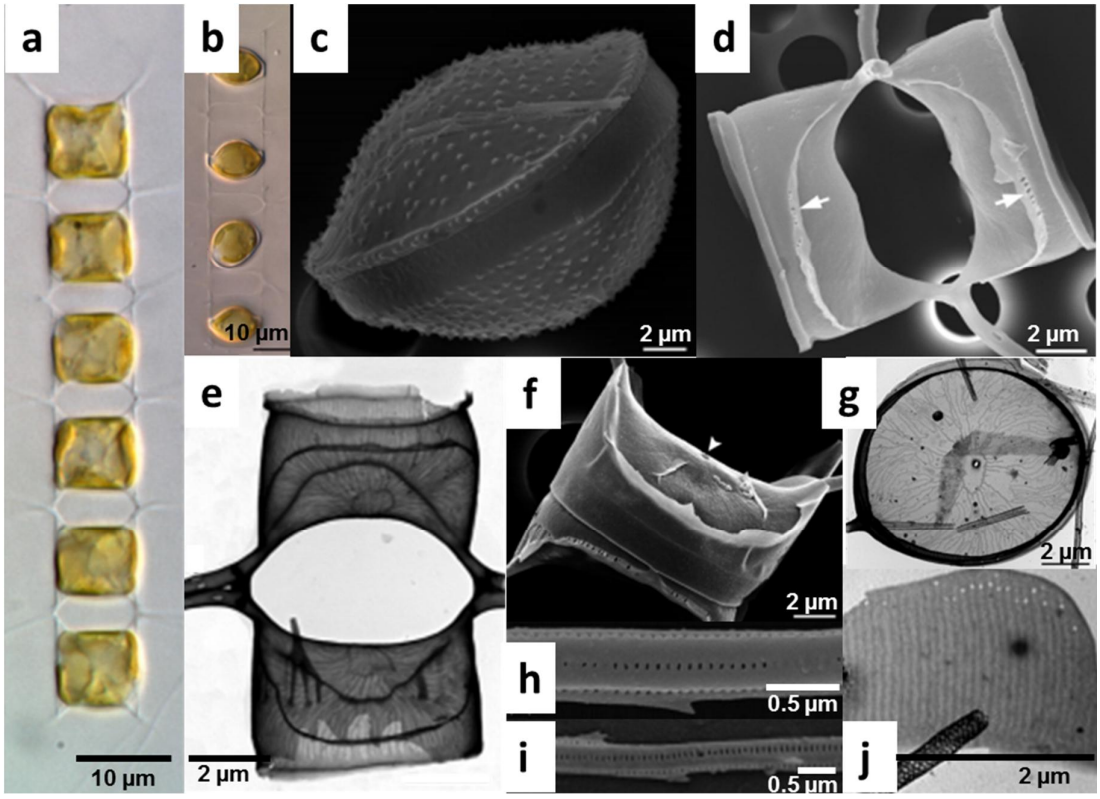


Fig. 3.15. *Chaetoceros dayaensis*. Strain MC107L and MC107S Figs (a−j) are reproduced from Li *et al.* (2015). LM (a−b), SEM (c−d, f, h and i) and TEM (e, g, and j). (a) A colony in girdle view. (b) Resting spores in mother cells. (c) Resting spore. (d-e) Sibling intercalary valves showing peanut-shaped aperture (arrow mark row of holes). (f−g) Terminal valve (arrow mark rimoportula). (h) Terminal seta. (i) Intercalary seta. (j) Girdle band.

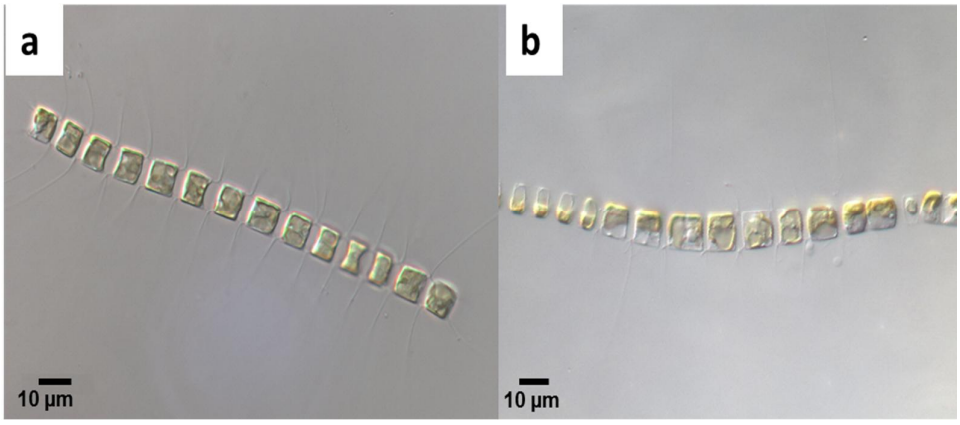


Fig. 3.16. *Chaetoceros* cf. *difficilis*. Strain Na2A3 Fig. (a) and Na26A2 Fig. (b). Figs (a-b). LM (a−b). (a−b) Colony in girdle view.

***C.* cf. *difficilis* Cleve**

Fig. 3.16. a−b

Eight stains highly similar in morphology were morphologically characterised as *C. cf. difficilis* based on an illustration in Hustedt (1930). Ultrastructural details of these strains were not examined. The described morphology corresponds to the strains Na2A3 (Fig. 3.16. a) and Na26A2 (Fig. 3.16. b). Chains long, slightly curved, and showing torsion (Figs 3.16. a−b). Cells quadrangular in girdle view and containing a single chloroplast. Apertures wide and hexagonal. Setae arise from the valve corners, cross over the chain axis and are directed perpendicular to chain axis. Terminal setae straight and oriented parallel to the axis of the colony.

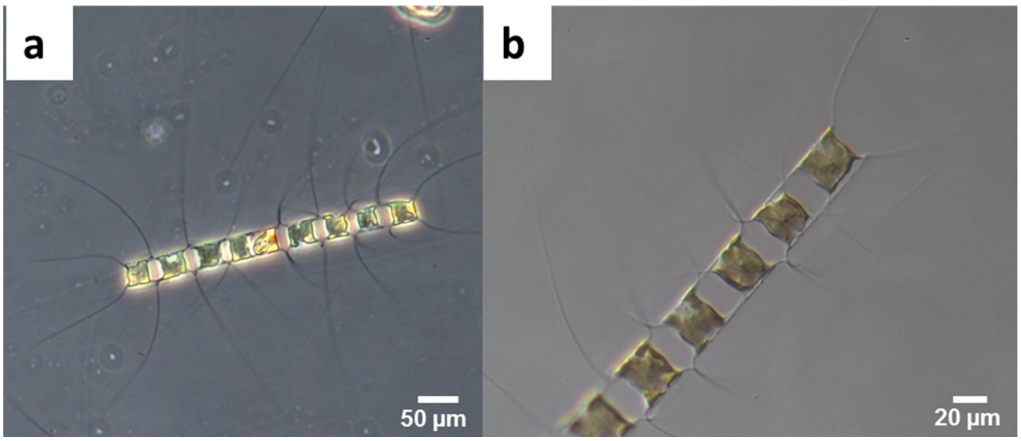


Fig. 3.17. *Chaetoceros* cf. *holsaticus*. Strain Na13C1 Figs (a−b). LM (a−b). (a−b) A colony in girdle view.

C. cf. *holsaticus* Schütt

Fig. 3.17. a−b

A single strain (Na13C1) was isolated during the study. It's gross morphology matched the description provided by Jensen & Moestrup (1998). Ultrastructural details of these strains were not examined in the present study hence the strain was referred as *C*. cf. *holsaticus*. Cells usually in straight chains, with wide hexagonal apertures. The valve face is concave, generally with a prominent central inflation (Figs 3.17. a−b). Each cell contains a single chloroplast. Mantle low or equidimensional with the girdle, a slight constriction is visible near the suture. Long intercalary setae originate from the corners of the valves, cross slightly outside the chain axis, and orient in all directions. Terminal setae straight, almost parallel to the axis of the colony and generally diverge from the apical plane giving a V-shaped appearance.

C. cf. *pseudocrinitus* Ostenfeld

Fig. 3.18. a−h

During this study, seven strains were isolated. Described morphology is based on strains Na12A3 and Na14A2. Chains straight or slightly bent and long (Figs 3.18. a−b). Cells elliptical to circular in valve view and rectangular in girdle view, with low mantle. Aperture narrow, slit-like, and covered with a silica jacket (Fig. 3.18. c). Each cell contains a single large chloroplast. Intercalary setae thin compared to the pronounced V-shaped terminal setae (Figs 3.18. a−b). Setae arise from the valve corners and cross immediately in the chain axis (Figs 3.18. c and f). Intercalary setae circular in cross-section, ornamented with spirally arranged shark-fin spines and minute poroids, with equidistantly arranged large pores (Figs 3.18. g−h). Terminal setae straight, almost parallel to the axis of the colony and generally diverge from the apical plane, appearing broadly V-shaped. Terminal valve with a central

slit-like rimoportula, internally (Fig. 3.18. d) and a short, flat external process of the rimoportula (Fig. 3.18. e). Resting spores not observed in the study.

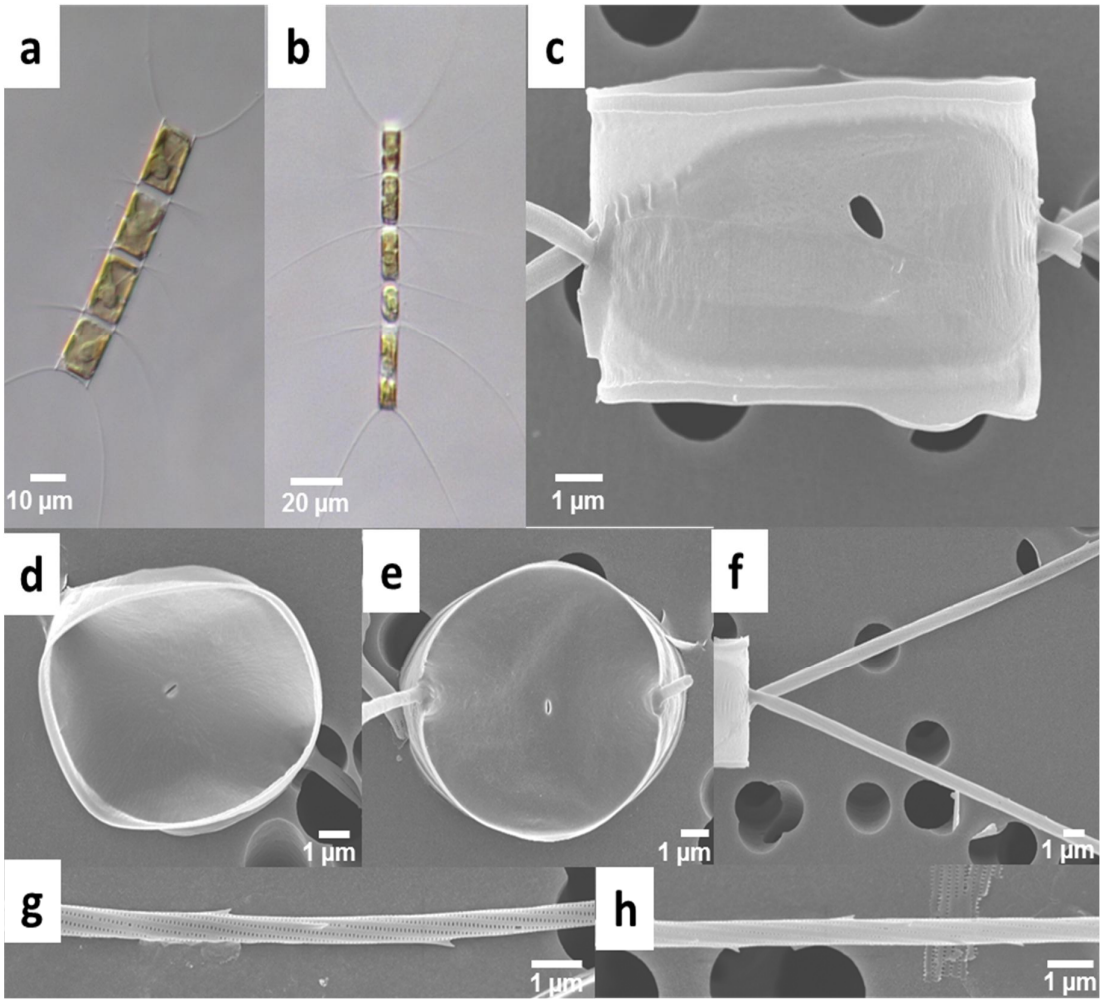


Fig. 3.18. *Chaetoceros* cf. *pseudocrinitus*. Strain Na12A3 Fig. (b) and Na14A2 Figs (a and c–h). LM (a–b) and SEM (c–h). (a–b) Colony in girdle view. (c) Intercalary valve, note the silica jacket covering the aperture. (d–e) (d) Internal view of a terminal valve with a central slit-shaped rimoportula. (e) External view of a terminal valve. (f) Detail of setae base. (g–h) Details of intercalary setae.

The gross morphology of these isolated strains was similar to that reported in Jensen & Moestrup (1998) and Shevchenko *et al.* (2006). However, the ultrastructural details do not correspond. Jensen & Moestrup (1998) illustrated the terminal valve with a central process with two outgrowths at the setae base (not observed in the present study). Shevchenko *et al.* (2006) illustrated the terminal valve with eccentric rimoportula. Because of this discrepancy, these strains were named as *C.* cf. *pseudocrinitus*.

***C. cinctus* Gran**

A detailed description of this species is presented in Chapter 5.

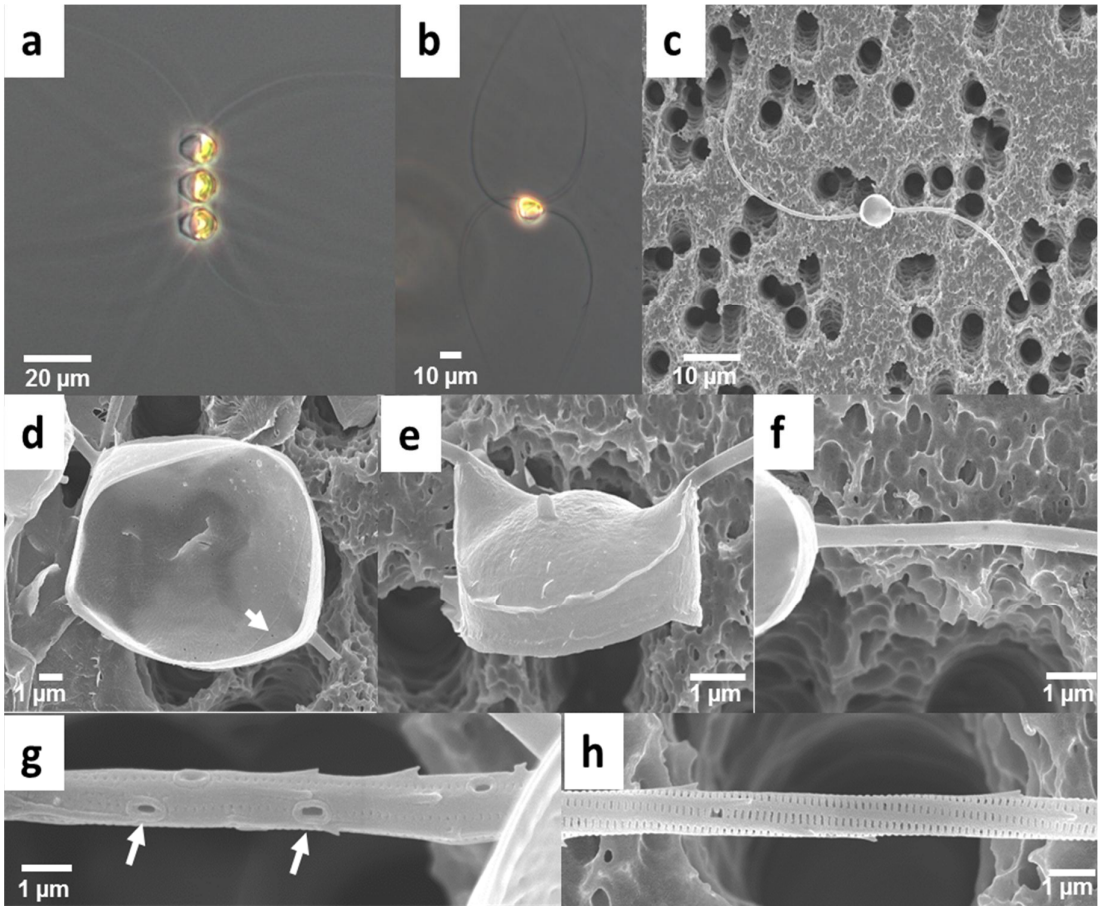


Fig. 3.19. *Chaetoceros circinalis*. Strain Na15C2 Figs (a–h). LM (a–b) and SEM (c–h). (a) A colony in girdle view (b) Solitary cell. (c) Terminal valve with S-shaped setae orientation. (d) Internal view of a terminal valve with a central slit-shaped rimoportula (arrow mark poroid). (e) External view of a terminal valve with tube-like process of the rimoportula. (f) Detail of seta base. (g) Detail of seta (arrow mark large solitary pore). (h) Detail of a seta with elongated poroids.

***C. circinalis* (Meunier) Jensen & Moestrup**

Fig. 3.19. a–h

Only one strain (Na15C2) was isolated during the study. The morphology of this strain matched the description provided by Jensen & Moestrup (1998). Cells usually united in short curved chains (Fig. 3.19. a), solitary cells were seen in culture (Fig. 3.19. b). Cells cylindrical, elliptical in valve view and rectangular in girdle view. One chloroplast per cell. Setae arise from the valve corners and cross on the chain axis. Apertures narrow. Intercalary

setae markedly curved (Figs 3.19. b−c). Setae circular in cross sections, with longitudinal slits extended parallel to setae axis, and spirally arranged shark-fin spines (Figs 3.19. g−h). Large pores present equidistantly along the setae (Fig. 3.19. g). Valve face slightly convex and densely perforated with numerous irregularly shaped poroids. (Fig. 3.19. d). Terminal valve with a central slit-like rimoportula (Fig. 3.19. d) which often has a tube-like process externally (Fig. 3.19. e). Resting spores not observed.

C. contortus **Schütt**

Fig. 3.20. a−h

The examined morphology corresponds to the strain Ch8A1. Chains often long, slightly twisted, with wide apertures between adjacent cells (Fig. 3.20. a). Multiple chloroplasts per cell. Cells usually wider than longer with rounded corners. Setae arise from inside of valve margins with a long basal part, and cross on chain axis. Intercalary setae of two types, common and specialised (Figs 3.20. a−b). Specialised intercalary setae thicker compared to normal setae, located towards one end of the colony, and possess large spines arranged in a spiral pattern (Fig. 3.20. f). Normal intercalary setae circular in cross section, ornamented with arrowhead-like spines, and with S-shaped slits (Fig. 3.20. g). Terminal setae show a similar pattern as normal intercalary setae. Valve broadly elliptical to circular (Fig. 3.20. c). Terminal valve with a central slit-shaped rimoportula (Fig. 3.20. d), externally appear as a flattened tube (Fig. 3.20. e). Girdle bands ornamented with transverse parallel costae with minute scattered poroids (Fig. 3.20. h). The specialised intercalary setae were observed only in the Chilean field material but were not formed when grown in culture. Resting spores were not seen in the cultured Chilean material.

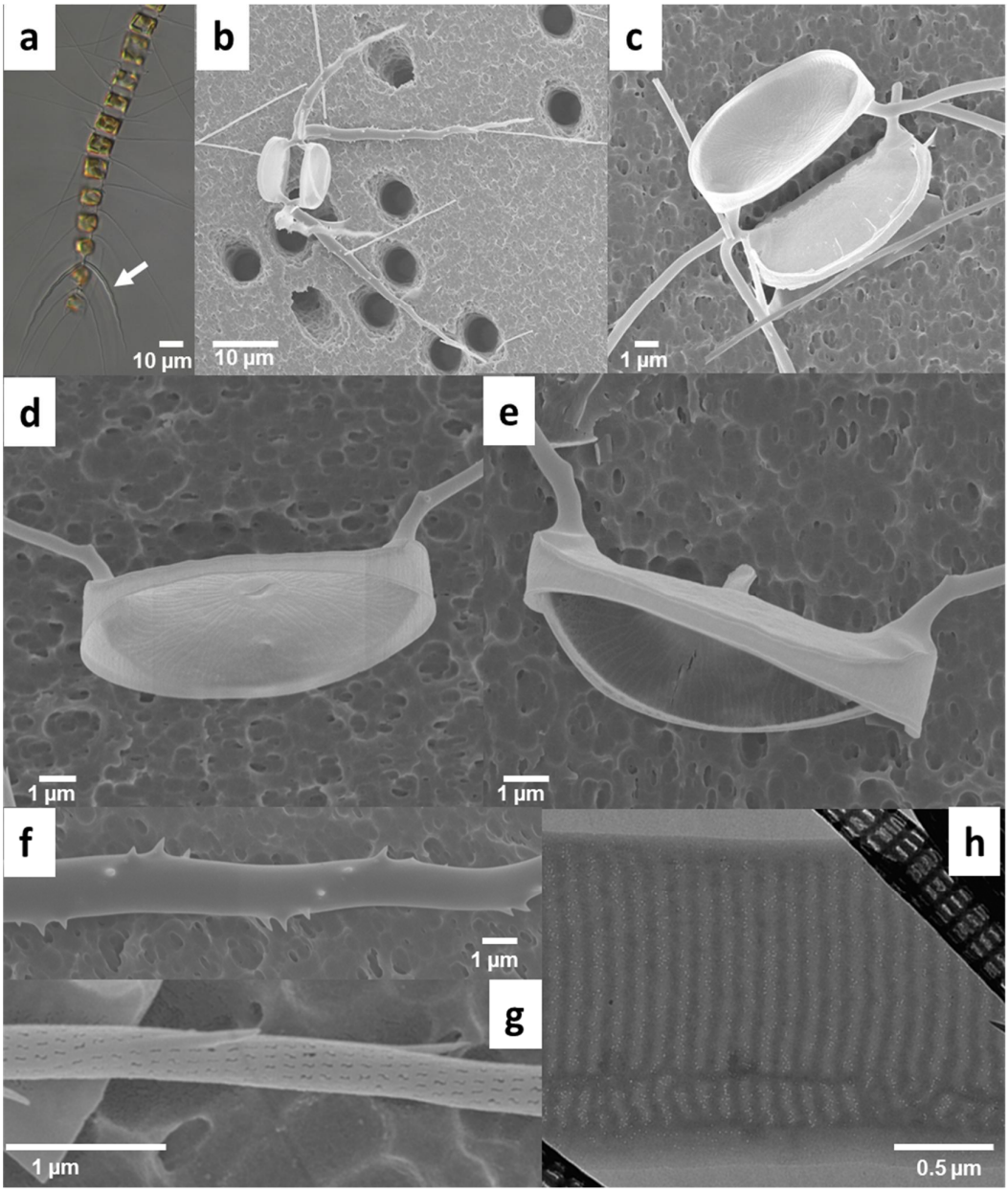


Fig. 3.20. *Chaetoceros contortus*. Strain Ch8A1 Figs (a−h). LM (a), SEM (c−g) and TEM (h). (a) A colony in girdle view (arrow mark specialised seta). (b) Intercalary valves with specialised setae. (c) Intercalary valves with normal setae. (d) Internal view of a terminal valve with the central slit-shaped rimoportula. (e) Terminal valve with the external tube-shaped process. (f) Detail of specialised seta. (g) Detail of normal seta, note the S-shaped slits on seta. (h) Girdle band.

Several varieties have been reported in the literature, which differ in resting spore morphology, setae morphology, spine type, the number of poroids and spines per length of seta, capilli on setae, specialised setae, and rimoportula number (Rines 1999; Lee *et al.* 2014a, 2014b; Chamnansinp *et al.* 2015). The Chilean strain revealed setae ultrastructural details -with S-shaped longitudinal slots, parallel to the setae axis- that differed slightly from

the ones reported in the literature (Kooistra *et al.* 2010; Chamnansinp *et al.* 2015). Several strains were isolated from the Gulf of Naples, but a 28S rDNA sequence available for only one of them (Na31B2), was identical to that of strain SZN-B402 from Kooistra *et al.* (2010). The latter strain was identified subsequently as *C. contortus* cf. *var. contortus* by Chamnansinp *et al.* (2015). The Chilean strains differed genetically from the Neapolitan strains and from the other varieties described by Chamnansinp *et al.* (2015).

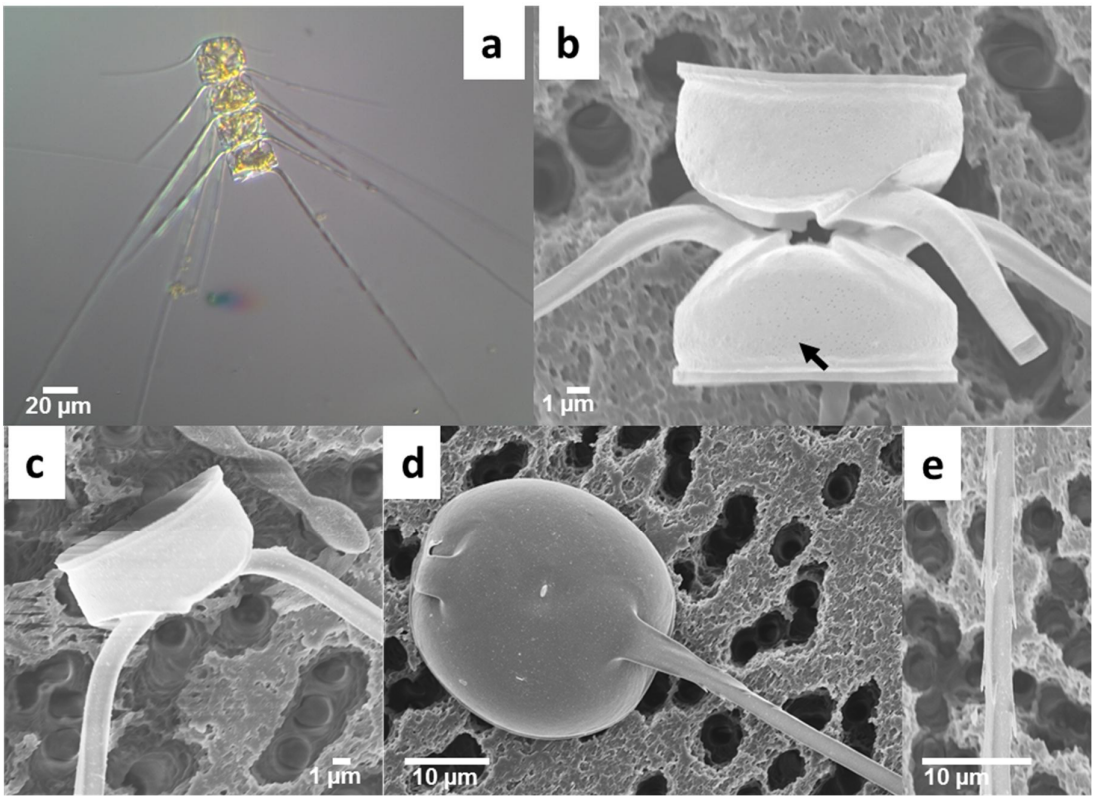


Fig. 3.21. *Chaetoceros convolutus*. Strain Ch5C3 Figs (a−e). LM (a), SEM (b−e). (a) A colony in girdle view (b) Intercalary valve (arrow mark poroid on valve). (c) Valve showing setae orientation. (d) Terminal valve with central rimoportula. (e) Detail of seta.

C. convolutes **Castracane**

Fig. 3.21. a−e

Two strains were identified as *C. convolutes* in this study. The morphology of strain Ch5C3 was examined and matched the description provided by Jensen & Moestrup (1998). Cells usually forming short chains, sometimes twisted, apertures very narrow, almost absent (Fig. 3.21. a). Several globular chloroplasts, extending in setae. Cells elliptical in valve view and

quadrangular in girdle view, with rounded apices. Mantle high. The setae arise from inside the valve corners (Fig. 3.21. c), bend towards the posterior end of the colony. Setae square in cross section (Fig. 3.21. b) with thick spines (Fig. 3.21. e). All valves ornamented with several poroids (Fig. 3.21. b) and possess a central rimoportula with a small tube-like external projection (Fig. 3.21. d). Resting spores not observed.

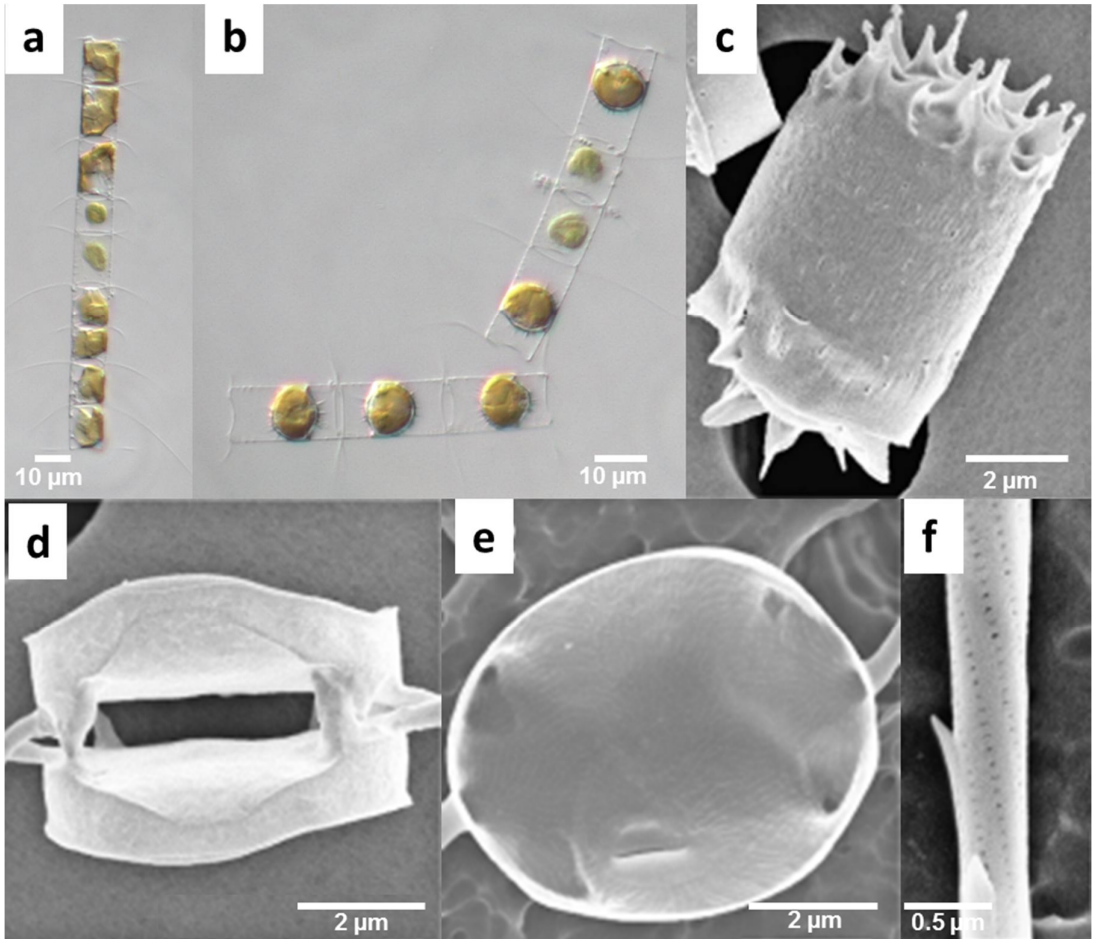


Fig. 3.22. *Chaetoceros costatus*. Strain Na7A4 Figs (a−b); figs (b−f) are reproduced from Kooistra *et al.* (2010). Figs (a−f). LM (a−b) and SEM (c−f). (a) A colony in girdle view (b) Resting spores in mother cells. (c) Resting spore. (d) Aperture. (e) Terminal valve with an eccentric slit-shaped process. (f) Detail of a seta.

***C. costatus* Pavillard**

Fig. 3.22. a−f

During this study nine strains corresponding to *C. costatus* were collected, all of these were genetically identical to the ones reported in Kooistra *et al*. (2010). The morphology of strain

Na7A4 was examined and corresponded to that reported in Kooistra *et al.* (2010). Cells forming long chains, usually straight or bend with single chloroplast (Fig. 3.22. a). Each valve possesses a sub-marginal protuberance close to valve margin that joins with adjacent cells, thus forming lanceolate apertures (Fig. 3.22. d). Intercalary setae emerge from the valve corner without a basal part, diverge from the chain axis and remain perpendicular to chain axis. Terminal setae undifferentiated from intercalary setae. Setae circular in cross section with spirally arranged spines and rows of poroids (Fig. 3.22. f). Valves broadly elliptical to circular. Terminal valve possesses an eccentric slit-shaped process (Fig. 3.22. e). Resting spores rounded with dome shaped appearance with numerous long spine visible in LM (Figs 3.22. b−c). SEM observations revealed that the primary valve is covered with hooked spines joined at their basal part by siliceous vela and the secondary valve has conical spines and flaps of silica.

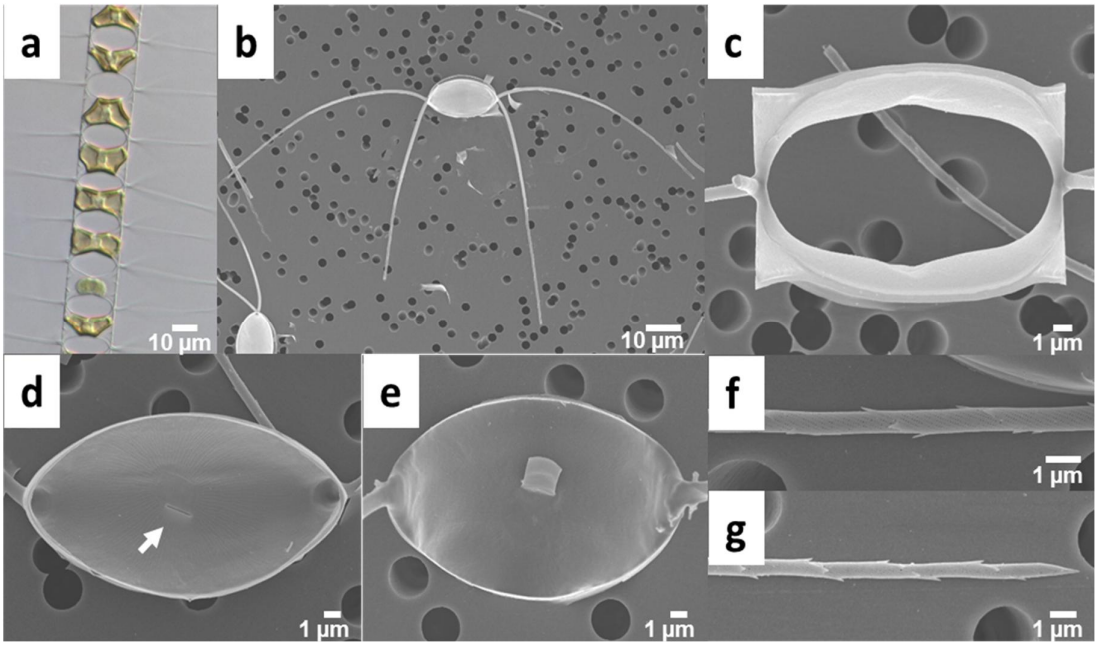


Fig. 3.23. *Chaetoceros curvisetus*. Strain Na10B4 Figs (a−j). LM (a) and SEM (b−g). (a) A colony in girdle view (b) Intercalary valve, note setae directed towards the side of the valve. (c) Aperture. (d) Internal view of the terminal valve with the central slit-shaped rimoportula (arrow mark slight eccentric process in central annulus). (e) External view of the terminal valve with the long tube-like rimoportula process (f) Detail of the middle part of the seta. (g) Tip of a seta.

***C. curvisetus* Cleve**

Fig. 3.23. a−g

The morphological details described her refer to strain Na10B4. Cells united to form curved chains, with large elliptical or circular aperture (Fig. 3.23. a). Each cell possesses one large chloroplast. All setae bend towards the convex side of the chain (Fig. 3.23. b). Intercalary setae emerge from the valve corners and cross on the chain axis. Terminal setae remain undifferentiated. Setae ornamented with a spiral pattern of shark-fin spines and rows of poroids (Figs 3.23. f−g). The valve face is elliptical to circular. Terminal valve have an eccentric elongated slit-shaped rimoportula with long external tube structure (Fig. 3.23. d, e). Resting spores not observed for this strain.

Fifteen strains were collected and identified as *C. curvisetus* during this study. Their 28S sequences grouped in four distinct clades within a major clade that also contained *C. pseudocurvisetus.* These clades have been named *C. curvisetus* 1, 2, 3, and 4 (which only contained the Chilean strains). Two of the Neapolitan clades (*C. curvisetus* 1, 2) were already reported in Kooistra *et al.* (2010). However, in the 18S phylogeny, the Chilean clade and one of the Neapolitan clades (*C. curvisetus* 1) were recovered closely together. Ultrastructural differences among vegetative cells of strains belonging to the different clades were not observed. Resting spore morphology was described only for *C. curvisetus* 1 (SZN-B406) in Kooistra *et al.* (2010; figs 43−45, 47), with smooth and round spores. Both valves possess a silica collar, which is higher in the primary valve and perforated with holes or fissures.

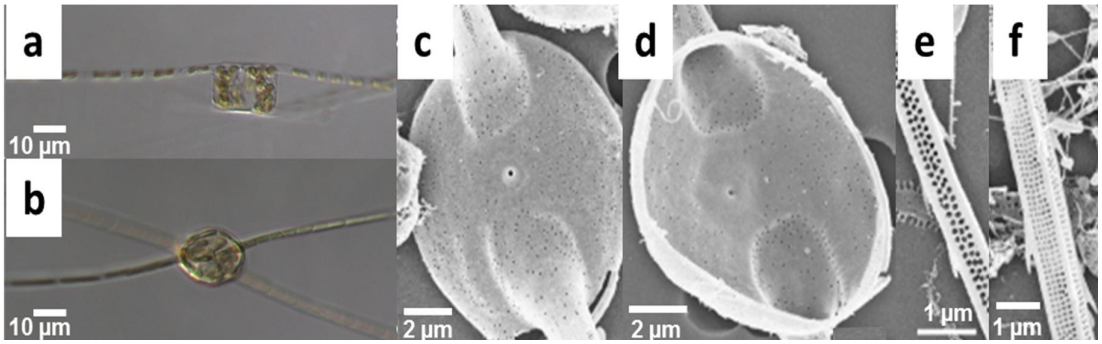


Fig. 3.24. *Chaetoceros danicus*. Strain Na9B4 Figs (a−b) and Figs (c−f) are reproduced from Kooistra *et al.* (2010). LM (a−b) and SEM (c−f). (a) A cell in girdle view. (b) A cell in valve view. (c) External view of valve, note numerous poroids on valve surface. (d) Internal view of valve, with the hole-shaped rimoportula. (e−f) Detailed structure of setae.

***C. danicus* Cleve**

Fig. 3.24. a−f

Only one strain (Na9B4) was isolated; the ultrastructure was not studied. The LM morphology and the 28S sequence matched with what reported by Kooistra *et al.* (2010), on which the illustration of the ultrastructure is based. Cells occur mostly solitary (Fig. 3.24. a), but short chains of two of three cells are, at times, observed. Several chloroplasts in the cell body as well as in the setae. Setae long and straight, arising from the valve corners, and then oriented perpendicular to chain axis. The setae diverging from the apical axis by ca. 15° (Fig. 3.24. b). Setae quadrangular in cross sections, with rows of spines on the edges. Each side of the setae possesses three to five single longitudinal row of large pores (Figs 3.24. e−f). Valves perforated with numerous pores (Figs 3.24. c−d). Each valve possesses a central hole-shaped rimoportula lacking an external tube (Figs 3.24. c−d). Resting spores not reported.

***C. debilis* Cleve**

Fig. 3.25. a−h

The morphological details described her refers to strain Ch13A4. Cells form spiral chains, with wide hexagonal apertures between adjacent cells (Figs 3.25. a−b), each cell containing a single chloroplast. Intercalary setae long and emerge from the corners of the valve and

cross just outside the chain axis (Fig. 3.25. f), and bend towards the convex side of the colony (Fig. 3.25. c). Setae ornamented with a spiral pattern of shark-fin spines and row of poroids, with solitary large pores arranged equidistantly (Figs 3.25. g–h). Valve face elliptical to circular, ornamented with radially branched costae, branching from an eccentric annulus. The terminal valves possess an elongated slit-shaped rimoportula (Fig. 3.25. d), which has a flattened tube externally (Fig. 3.25. e). Resting spores not seen in the culture material.

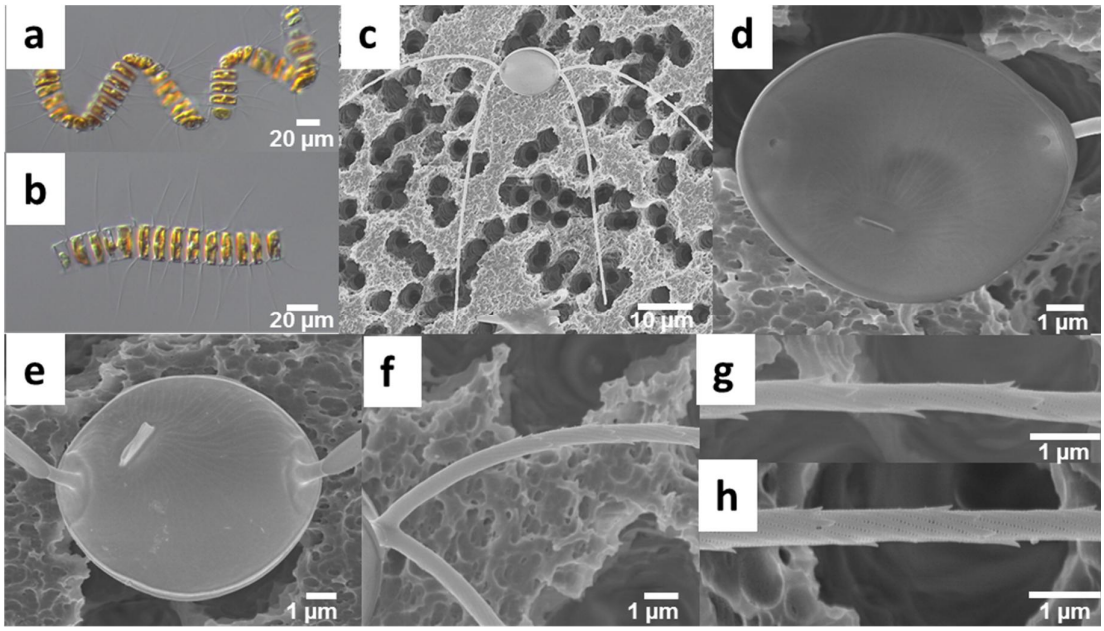


Fig. 3.25. *Chaetoceros debilis*. Strain Ch13A4 Figs (a–h). LM (a–b) and SEM (c–h). (a–b) Colonies in girdle view. (c) Intercalary valve, note setae directed towards the side of the valve. (d) Internal view of terminal valve with the eccentric slit-shaped rimoportula. (e) External view of terminal valve with the flattened process (f) Detail of setae base. (g–h) Details of setae.

Six strains were isolated from Chilean waters during the study and the gross morphology of vegetative cells matched the description provided by Kooistra *et al.* (2010). Kooistra *et al.* (2010) reported two genetically distinct but morphologically identical species, one represented by strains from Helgoland, and one by strains from the Pacific Ocean. The Chilean strains form a third clade. An 18S sequence retrieved from GenBank constitutes a fourth lineage. However, spores were also not detected by Kooistra *et al.* (2010). Paired spores are however reported for *C. debilis* (Rines & Hargraves 1988; Jensen & Moestrup

1998) and a detailed investigation of the spore ultrastructure for the two genetically different groups is required.

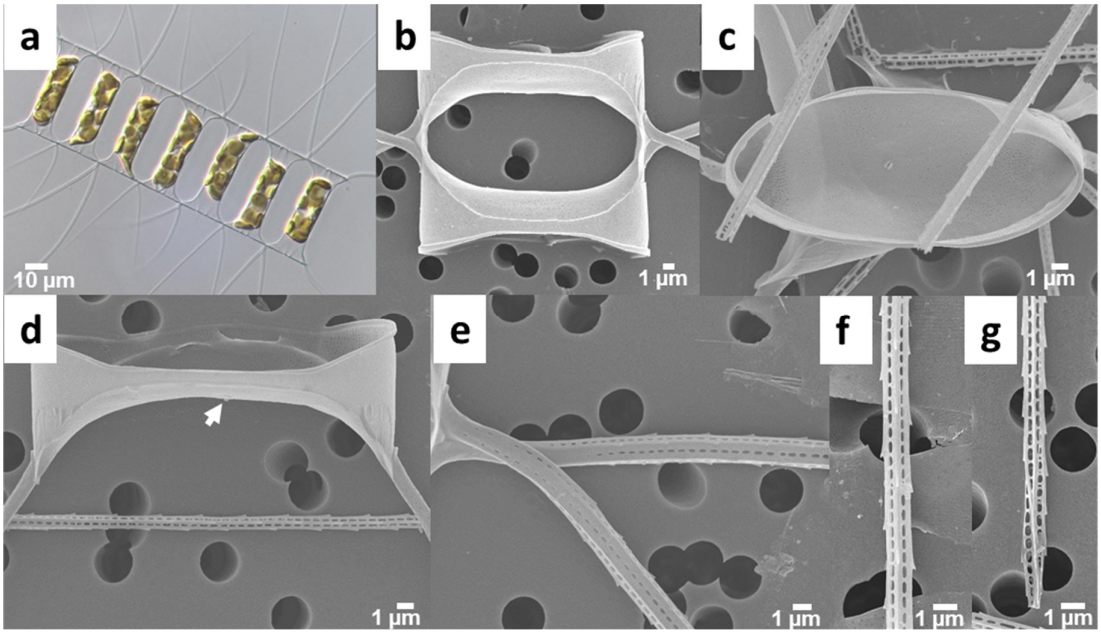


Fig. 3.26. *Chaetoceros decipiens*. Strain Ro2C4 Fig. (a) and Na11B3 Figs (b−g). LM (a) and SEM (b−g). (a) A colony in girdle view (b) Two sibling valves. Note the wide aperture. (c) Terminal valve, note numerous poroids. (d) Internal view of the terminal valve with central slit-shaped rimoportula. (e) External view of the terminal valve with the short rimoportula process (arrow mark rimoportula). (e) Detail of setae base. (f) Detail of middle part of seta. (g) Tip of a seta.

C. decipiens **Cleve**

Fig. 3.26. a−g

During this study fifteen strains corresponding to *C. decipiens* were collected. The morphological illustration is based on strains Ro2C4 and Na11B3. The morphology of this strain matched the description provided by Jensen & Moestrup (1998) and Kooistra *et al.* (2010, as *C. lorenzianus*). Cells forming short chains, with wide, elliptical or oval apertures between adjacent cells (Fig. 3.26. a-b). Each cell contains multiple chloroplasts. Intercalary setae emerge on the valve apices with a short basal part, fuse on the chain margin and positioned in the apical plane (Fig. e). Terminal setae straight, almost parallel to the axis of the colony forming a U-shape. Setae four-sided in cross section and possessed numerous longitudinally arranged large pores running parallel to setae axis and spirally arranged spines

on edges (Fig. 3.26. f−g). Valves elliptical in outline, perforated with numerous pores (Fig. 3.26. c). Terminal valves possess a central small slit-like rimoportula (Fig. 3.26. c) with a short external process (Fig. 3.26. d). Resting spores were not seen in the culture material.

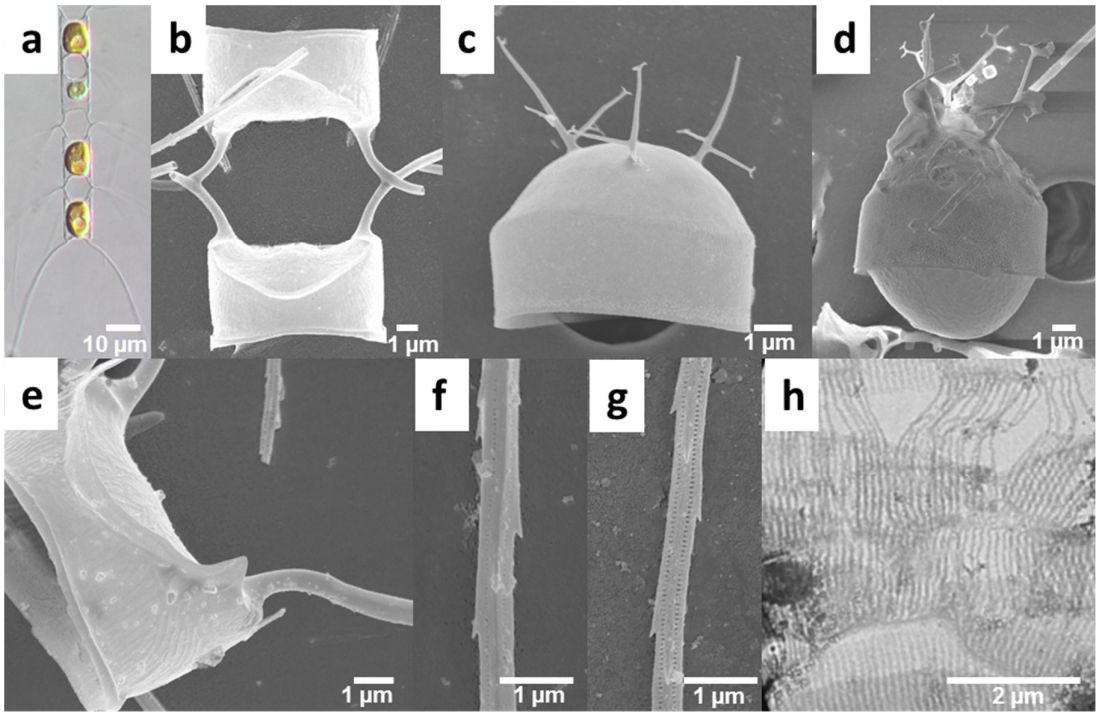


Fig. 3.27. *Chaetoceros diadema*. Strain Ch1C1 Fig. (a) and Ch4A1 Figs (b−h). LM (a), SEM (b−g) and TEM (h). (a) A colony in girdle view (b) Sibling valves. Note the hexagonal aperture. (c) Resting spore, primary valve with incompletely formed spines. (d) Resting spore. (e) Terminal valve with the central external tube of the rimoportula. (f) Detail of the terminal seta. (g) Detail of intercalary seta. (h) Girdle bands.

***C. diadema* (Ehrenberg) Gran**

Fig. 3.27. a−h

In the present study, thirteen strains from Chile and Naples were isolated. The description is based on strains Ch1C1 and Ch4A1. Chains short, with wide hexagonal apertures between adjacent cells (Fig. 3.27. a). Each cell contains a single chloroplast. Intercalary setae emerge on the valve corner with a conspicuous basal part. They cross just on the chain margin and point in all directions (Fig. 3.27. b). Terminal setae straight, almost parallel to the axis of the colony and generally diverged from the apical plane. Setae quadrangular in cross section and possess numerous longitudinally arranged poroids (Figs 3.27. f−g). Spines arranged in a

spiral pattern. Valve elliptical in outline, ornamented with radially branched costae. Valve face adorned with small capilli (Fig. 3.27. e). Terminal valves possess a central rimoportula with a small external process (Fig. 3.27. e). Terminal valves show a horn at the base of the setae emergence (Fig. 3.27. e). Resting spores composed of two vaulted valves, with dichotomously branched spines on the primary valve (Figs 3.27. c−d).

The morphology of all strains matched with the descriptions of *C. diadema* provided by Jensen & Moestrup (1998) and Kooistra *et al.* (2010). However, the strains isolated in this study, cluster into three different lineages, one with Neapolitan, and the other two from Chilean strains.

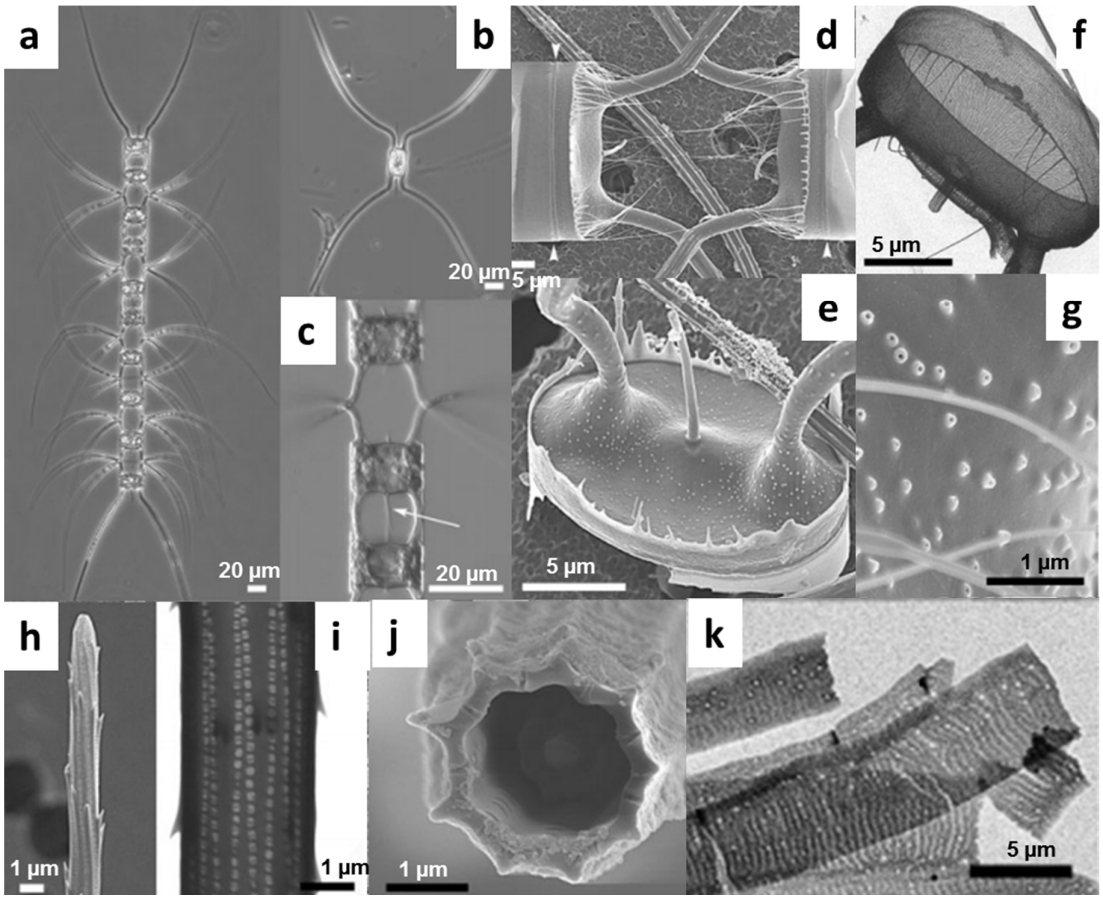


Fig. 3.28. *Chaetoceros dichaeta*. Strain MM19-D3 Figs (a−k) are reproduced from Assmy *et al.* (2008). LM (a−c), SEM (d−e, g−h, and j) and TEM (f, i and k). (a) A colony in girdle view (b) Solitary cell. (c) Cells characterized by a short pervalvar and a wide apical axis (arrow mark long tubular process on the intercalary valve. (d) Aperture. (e−f) Terminal valve. (g) Poroids on the valve face. (h) Tip of a seta. (i) Detail of a seta. (j) Cross section of a seta. (k) Girdle bands.

C. dichaeta Ehrenberg

Fig. 3.28. a−k

This species was not collected in the present study, but DNA samples collected within the study of Assmy *et al.* (2008) was used to obtain the 18S sequences. The description of the species reported here is based on Assmy *et al.* (2008). Colonies in straight chains (Figs 3.28. a−c), solitary cells also reported. Aperture wide and rectangular (Fig. 3.28. d). Multiple chloroplasts distributed in the cell body and setae. Intercalary setae emerge slightly inside the valve margin with the long basal part and cross on the chain margin (Figs 3.28. c−d). Intercalary setae circular in cross-section but become polygonal after the junction. Setae had a longitudinal row of poroids and spirally arranged spines (Figs 3.28. h−i). Valves possess a long tubular process (Figs 3.28. e−f) and are ornamented with an irregular pattern of poroids (Fig. 3.28. g) and capilli on the valve margins (Fig. 3.28. d). Resting spores were not seen.

C. dichatoensis sp. nov.

A description of this new species is reported in Chapter 5.

C. didymus Ehrenberg

Fig. 3.29. a−h

Four strains were isolated in the present study. Illustrated morphology corresponds to strain Ch2B4 that matched the description provided by Shevchenko *et al.* (2006). Chains straight, with wide apertures between adjacent cells (Fig. 3.29. a). Each cell has two chloroplasts. The valve possessed a characteristic central inflation that appears as a protuberance. Intercalary setae long and emerge on the valve apices and cross just outside the chain margin, and at times curved towards the terminal part of the colony. Terminal setae straight, almost parallel to the axis of the colony and generally reveal a U-shape. Setae pentagonal in cross-section and possess numerous minute longitudinally arranged poroids and spirally arranged spines

(Figs 3.29. e–f). Valve elliptical in outline, with radially branched costae. Valve face possesses branched capilli (Figs 3.29. b–c). The terminal valve shows a slit-shaped rimoportula positioned in the centre of the protuberance (Fig. 3.29. d). Resting spores reported to occur in pairs but have not been observed in the isolated strains.

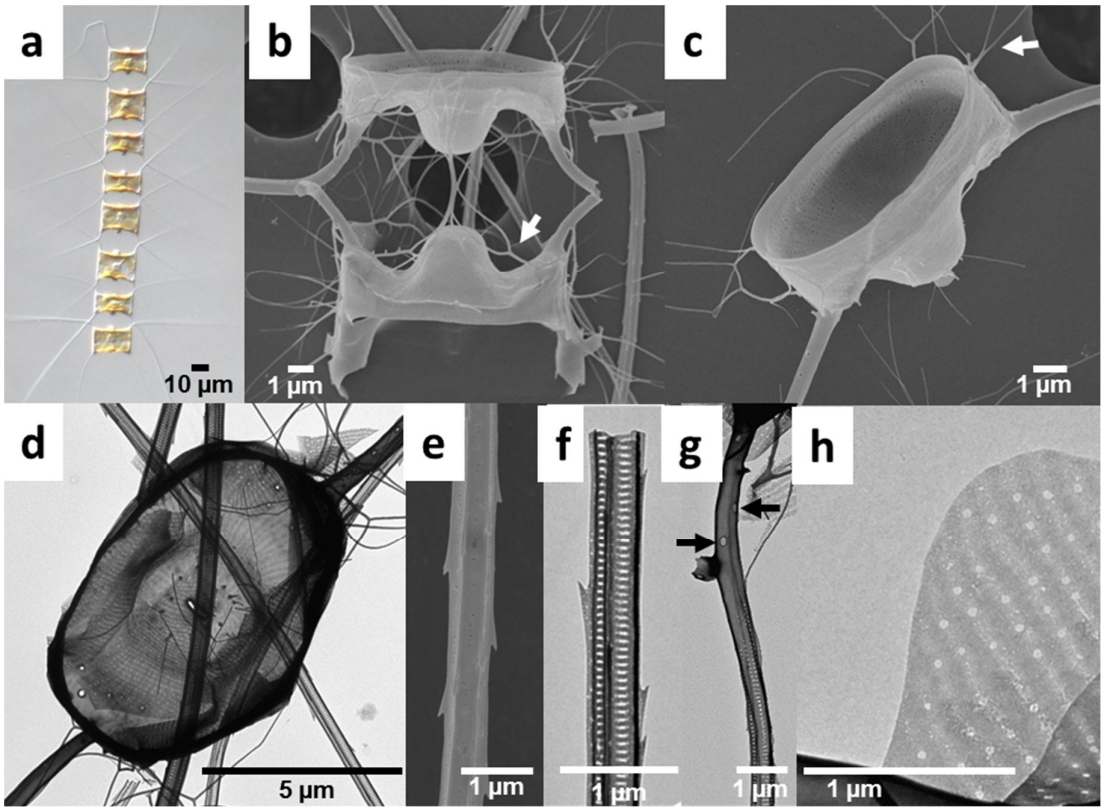


Fig. 3.29. *Chaetoceros didymus*. Strain Ch2B4 Figs (a–h). LM (a), SEM (b–c and e) and TEM (d and f–h). (a) A colony in girdle view (b) Sibling valves with central protuberances (arrow indicate capilli on valve face). (c–d) Terminal valve with a central slit-shaped process (arrow mark capilli on valve mantle). (e–f) Details of the middle part of setae. (g) Detail of seta at the base (arrows mark large pores on seta base). (h) Girdle band.

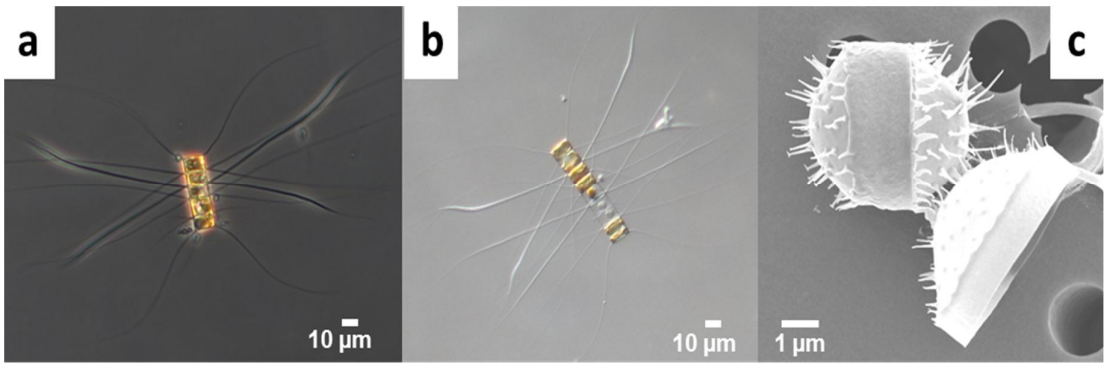


Fig. 3.30. *Chaetoceros diversus*. Strain Na3C1 Figs (a–c). LM (a–b) and SEM (c). (a–b) A colony in girdle view (b) Resting spore.

***C. diversus* Cleve**

Fig. 3.30. a−c

In this study, six strains were isolated, which correspond to the description of the species provided by Ruiz *et al.* (1993). Illustrated morphology corresponds to strain Na3C1. Cells tend to form straight chains, with a narrow and slit-shaped aperture between adjacent cells (Figs 3.30. a−b). Each cell contains a single central chloroplast. Cells elliptical in valve view and rectangular in girdle view. Setae originate from the valve corners, crossing immediately and diverging sideways, but remaining in the apical plane. Intercalary setae typically V-shaped (Figs 3.30. a−b). Two types of intercalary setae, the common setae and the thick specialized setae. Spores dome-shaped ornamented with conical spines (Fig. 3.30. c).

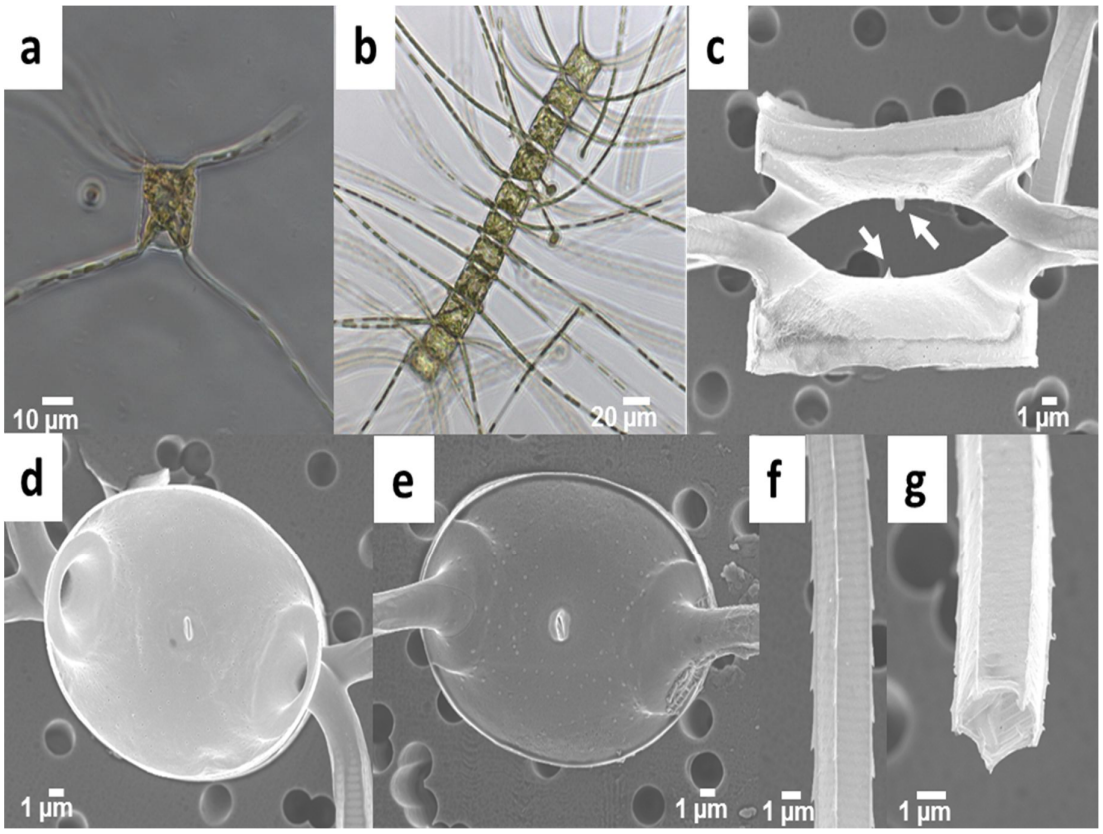


Fig. 3.31. *Chaetoceros eibenii*. Strain Ro1B2 Figs (a−g). LM (a−b) and, SEM (c−g). (a) Single cell.– (b) A colony in girdle view. (c) Sibling valves (arrows mark short tubular processes). (d) Internal view of the intercalary valve with a central slit-shaped rimoportula). (e) External view of the terminal valve with a central slit-shaped process, note poroids on the valve face. (f) Detail of seta. (g) Cross section of seta.

C. eibenii (Grunow in Van Heurck) Meunier

Fig. 3.31. a–g

Five strains were isolated, which matched the description provided by Jensen & Moestrup (1998). Illustrated morphology is of strain Ro1B2. Cells forming straight or slightly bend chains (Figs 3.31. a–b). Cells rectangular with rounded corners in broad girdle view. Aperture elliptical (Fig. 3.31. c). Numerous chloroplasts in the cell body and in the setae. Setae emerge near valve corners and cross on the chain margin (Fig. 3.31. c). Terminal setae oriented parallel to chain axis, forming a broad V-shape. Setae pentagonal in cross-section (Fig. 3.31. g), and show longitudinally arranged spine (Fig. 3.31. f). Valve perforated with large pores (Figs 3.31. d–e). Slit-shaped rimoportula with a short external tubular process present on each valve (Figs 3.31. c–e).

C. gelidus Chamnansinp, Li, Lundholm & Moestrup

A detailed description of this species is provided in Chapter 5.

C. lauderi Ralfs in Lauder

Fig. 3.32. a–j

Seven strains corresponding to *C. lauderi* were isolated. The partial 28S rDNA sequence and morphology of these strains are identical to those reported in Kooistra *et al.* (2010). Described morphology corresponds to strains Na1A1, Na13A4, and material in Kooistra *et al.* (2010). Cells united to form colonies (Figs 3.32. a–c). Aperture between adjacent cells narrow (Fig. 3.32. b). Multiple chloroplasts per cell. Intercalary setae long and emerge on the valve apices. They cross at the chain margin and directed in all directions (Fig. 3.32. h). Terminal setae straight, almost parallel to the axis of the colony. Setae circular in cross section and possess numerous spirally arranged poroids and spines (Fig. 3.32. j). Valves elliptical in outline, ornamented with radially branched costae. Terminal valves possess a

central slit-shaped rimoportula having a flattened tube-shaped process externally (Fig. 3.32. i). Resting spores present. The primary valve face is capitate with a constriction at its base and possesses thick spines near the centre and in a marginal ring while the secondary valve is ornamented with several long thin spines (capilli) (Figs 3.32. d–g).

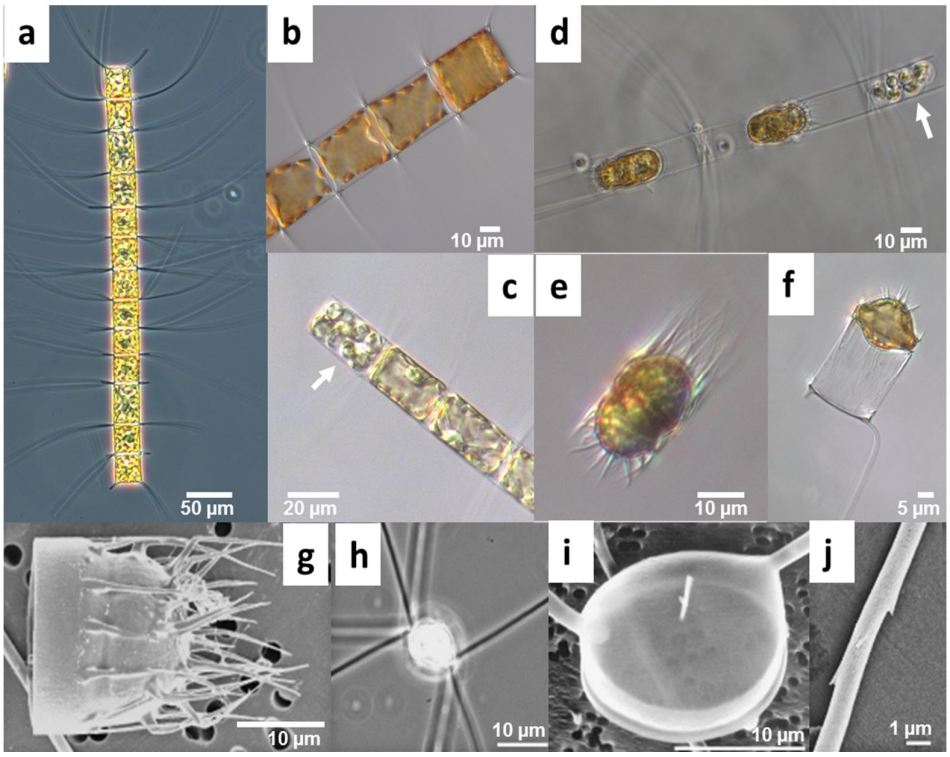


Fig. 3.32. *Chaetoceros lauderi*. Strain Na1A1 Figs (a–b, and f) and Na13A4 Figs (c–e) and Figs (g–j) are reproduced from Kooistra *et al.* (2010). LM (a-f) and SEM (g–j). (a–b) A colony in girdle view. (c) Part of a colony with a spermatogonium (arrowed). (d) Part of a colony in girdle view with two resting spores in the formation and a spermatogonium (arrow mark spermatogonium with sperm cells). (e–f) Resting spores. (g) Primary valve of resting spore. (h) Cells in valvar view. (i) Terminal valve with the central slit-shaped rimoportula. (j) Detail of a seta.

C. cf. *lorenzianus*

Fig. 3.33. a–k

Seventeen strains with morphology similar to *C. lorenzianus* were isolated from Chilean waters during this study. Illustrated morphology corresponds to the strain Ch12A1, whose gross morphology matched the description provided by Jensen & Moestrup (1998) and Sunesen *et al.* (2008). Cells forming short chains, with wide elliptical apertures between adjacent cells (Figs 3.33. a–b and d). Each cell contains multiple chloroplasts. Intercalary

setae emerge on the valve apices without a basal part, cross just on the chain margin and are positioned in the apical plane. Terminal setae straight, almost parallel to the axis of the colony forming a U-shape (Fig. 3.33. b). Setae pentagonal in cross-section and possess numerous longitudinally arranged large pores running parallel to setae axis and spirally arranged spines (Figs 3.33. i–j). Valves elliptical in outline, perforated with numerous pores (Fig. 3.33. e). Terminal valves with a small slit-shaped rimoportula (Fig. 3.33. g) bearing a short external process (Fig. 3.33. f). Resting spore with dissimilar valves (Figs 3.33. b–c). Primary valve with two apical and conical protuberances ornamented with dichotomously branching spines. Secondary valve smooth with one or two undulations.

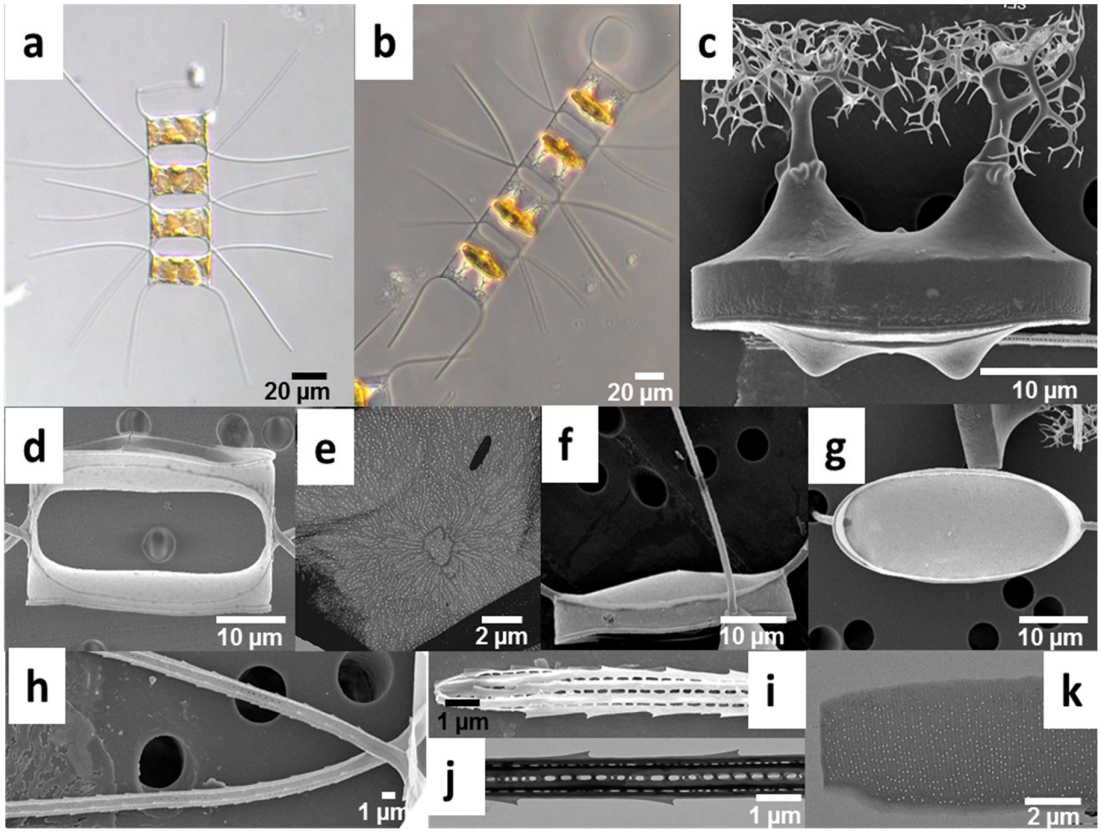


Fig. 3.33. *Chaetoceros* cf. *lorenzianus*. Strain Ch12A1 Figs (a–k). LM (a–b), SEM (c–d and f–i) and TEM (e and j–k). (a) A colony in girdle view (b) Colony with resting spores. (c) Resting spore. (d) Aperture. (e) Intercalary valve, note small scattered pores. (f) Terminal valve. (g) Terminal valve with a central slit-shaped process. (h) Detail of setae base. (i) Tip of a seta. (j) Detail of a seta. (k) Detail of a girdle band.

Based on the sequence information, three different genotypes were found (*C.* cf. *lorenzianus* 1, 2, 3). Strain Ch12A1 belongs to *C.* cf. *lorenzianus* 3, which matches the description by

Sunesen *et al.* (2008). *Chaetoceros* cf. *lorenzianus* 1 and *C.* cf. *lorenzianus 2* showed small pores arranged longitudinally on the setae; the aperture of *C.* cf. *lorenzianus* 1 is completely covered by a silica layer, which is absent in the other two genotypes. Spores were only observed for *C.* cf. *lorenzianus* 3.

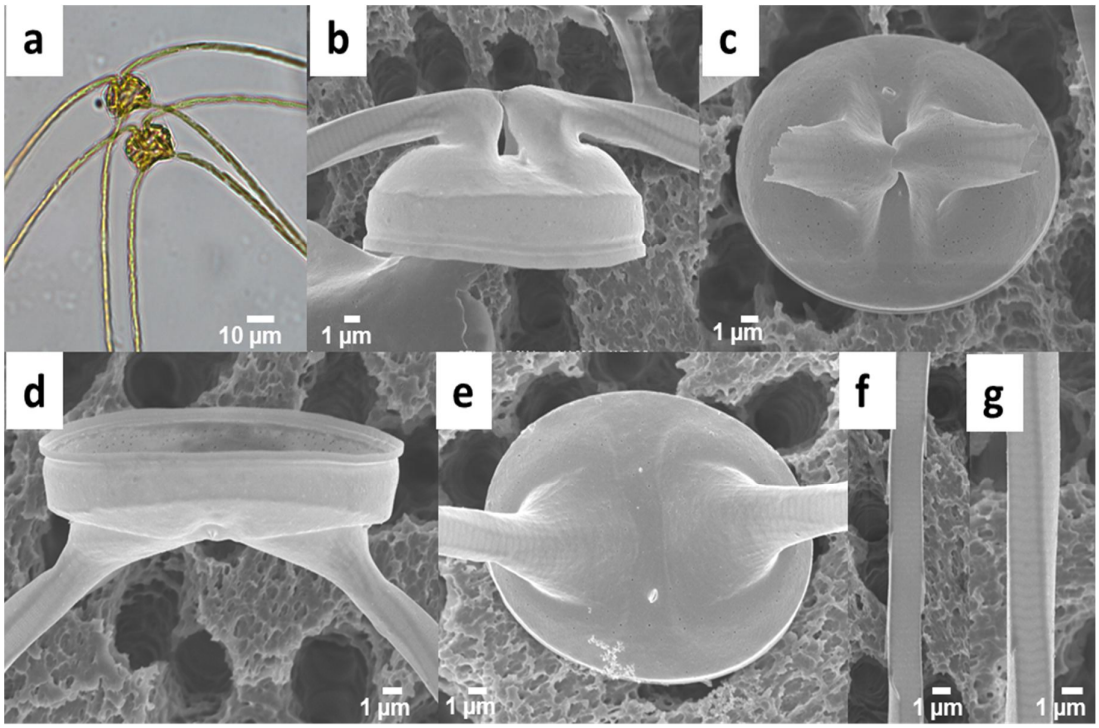


Fig. 3.34. *Chaetoceros peruvianus*. Strain Ch11B4 Figs (a−g). LM (a) and SEM (b−g). (a) Cells in girdle view. (b) Anterior valve in girdle view. (c) Anterior valve in valvar view, note presence of scattered poroids on the valve face. (d) Posterior valve in girdle view (e) posterior valve in valvar view. Note the slightly eccentric rimoportula. (f−g) Details of setae.

***C. peruvianus* Brightwell**

Fig. 3.34. a−g

Four strains were collected in this study. Illustrated morphology is of strain Ch11B4 that matched the description provided by Kooistra *et al.* (2010). Cells usually solitary, rarely in pairs (Fig. 3.34. a). Several small chloroplasts, located in the central body and in the setae. Cells elliptical in valve view and quadrangular in girdle view, with rounded apices (Fig. 3.34. b). The anterior setae arise from the valve centre, follow a pervalvar axis over a short distance and then fuse over the centre of the valve to reveal a tooth-like structure (Figs 3.34.

b−c), from where the setae bend towards the posterior end of the cell. The setae of the posterior valve direct them in a V-shape (Fig. 3.33. a). Setae square in cross section with spines and rectilinear grid-shaped pattern, with three transverse striae of poroids separated by a transverse costa (Figs 3.33. f−g). Valves ornamented with several poroids (Fig. 3.33. c), and possess a small tube like eccentric rimoportula with a short external projection (Figs 3.33. d−e). Resting spores not observed.

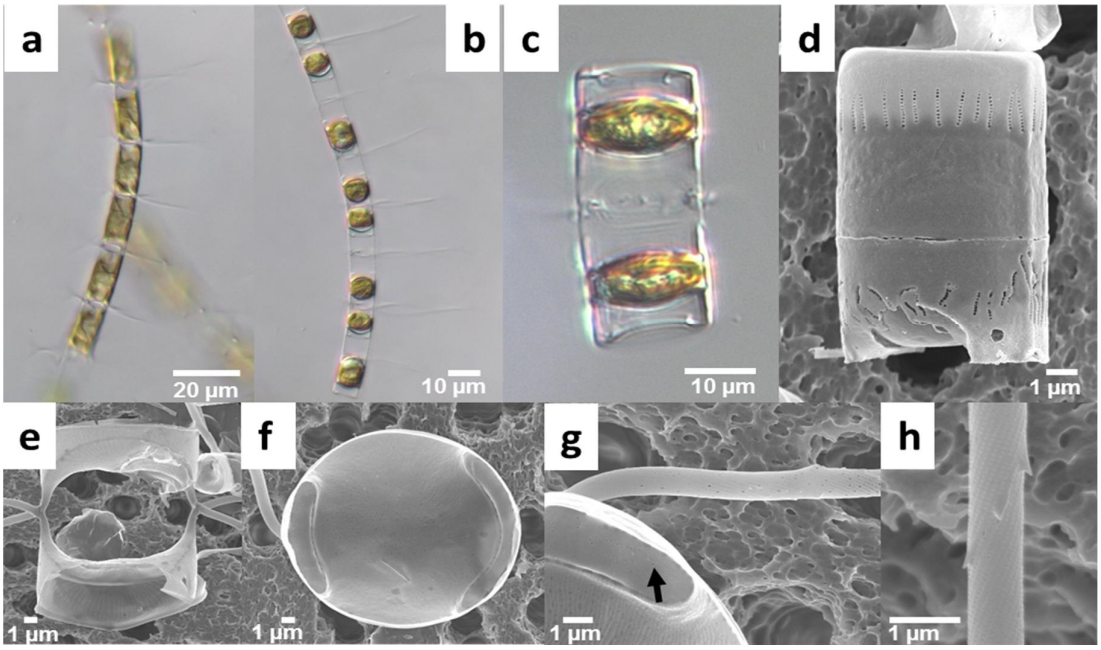


Fig. 3.35. *Chaetoceros pseudocurvisetus*. Strain Na13C4 Figs (a−h). LM (a−c), and SEM (d−h). (a) A colony in girdle view. (b) A colony with resting spores. (c−d) Two resting spores. (e) Aperture. (f) Terminal valve with an eccentric slit-shaped process. (g) Detailed structure of seta base, note the presence of randomly organised scattered large pores on setae base (arrow mark the presence of perforation on the valve face). (h) Detail of a seta.

***C. pseudocurvisetus* Mangin**

Fig. 3.35. a−h

In this study, two strain were collected, which match the description provided by Kooistra *et al.* (2010). Illustrated morphology is of the strain Na13C4. Cells united to form curved chains, with an elliptical or circular aperture (Figs 3.35. a−b). Each cell possesses one chloroplast. All setae bend towards the convex side of the chain. Intercalary setae emerge from the valve apices and cross on the chain axis as in *C. curvisetus* complex and *C. debilis*

complex. Each valve possesses four submarginal projections close to valve margins that join with adjacent cells, thus revealing lanceolate apertures (Fig. 3.35. e). Terminal setae undifferentiated. Setae ornamented with a spiral pattern of shark-fin spines and minute poroids (Fig. 3.35. h). Also, several large poroids scattered along the setae in between the minute pores and at the base of setae (Fig. 3.35. g). Valve face is elliptical to circular. Terminal valve reveals an eccentric annulus, with an elongated slit (Fig. 3.35. f). The resting spores of these species are smooth and rounded (Fig. 3.35. c), with both valves possessing a silica collar, which is higher in the primary valve and perforated with holes or fissures (Fig. 3.35. d).

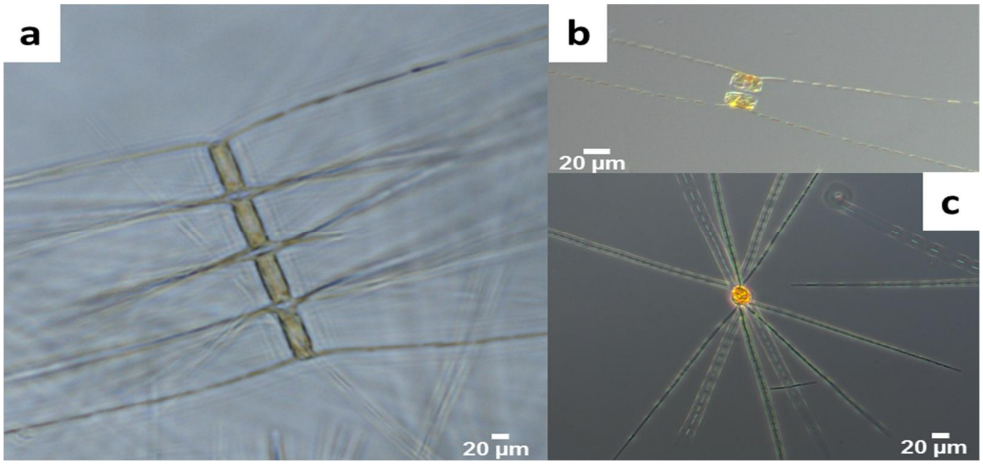


Fig. 3.36. *Chaetoceros rostratus*. Strain Na1B3 Figs (b−c) and Na6B3 Fig. (a). LM (a−c). (a) A colony in girdle view. (b) A cell in division. (c) A colony in valvar view.

C. radicans Schütt

The detailed description of this species is in Chapter 5.

C. rostratus Lauder

Fig. 3.36. a−c

During this study, four strains were isolated that revealed similar gross morphology as reported by Rines & Hargraves (1988). Illustrated morphology corresponds to strains Na1B3

and Na6B3; Information on the ultrastructure of these strains was not obtained. Colony straight and robust, with narrow apertures (Fig. 3.36. a). Solitary cells were often seen in culture (Fig. 3.36. b, illustrates a dividing cell). Cells joined in chains by a fusion of a typical central process present on the valve face and not by setae fusion. Numerous small chloroplasts spread in the cell as well as in setae. The setae arise from valve corners, and then diverge perpendicular to the chain axis, forming variable angles in the apical plane (Fig. 3.36. c). Terminal setae have the orientation of intercalary setae (Fig. 3.36. b). Resting spores not observed.

C. rotosporus Li, Lundholm & Moestrup

Fig. 3.37. a–g.

Ten strains of this morphotype were collected during this study. The 28S rDNA sequences of these strains are identical to the sequence of *C. rotosporus* from Li *et al.* (2013). The morphology of strain Na22A3, which is described here, confirms the detailed morphological information provided by Li *et al.* (2013).

Chains long and straight, with wide, and rounded quadrangular apertures between adjacent cells (Figs 3.37. a–b). Each cell contains a single chloroplast. Intercalary setae long and emerge on the valve margin with the long basal part and cross just outside the valve margin, and directed in all directions. Terminal setae straight, almost parallel to the axis of the colony and generally diverging from the apical plane giving a U-shape appearance (Figs 3.37. a–b). Setae four sided or quadragonal in cross section and possess four longitudinal rows of poroids and spine arranged alternately on the setae (Fig. 3.37. g). Valve mantle ornamented with parallel ribs (Fig. 3.37. f). Valve elliptical in outline, ornamented with radially branched costae (Fig. 3.37. d). Terminal valve with a central short external rimoportula (Fig. 3.37. e), and internally a small slit. Fully formed resting spores are rotated 90° from their original

orientation during formation and exhibit a silica wing (Fig. 3.37. c). Girdle bands show parallel costae.

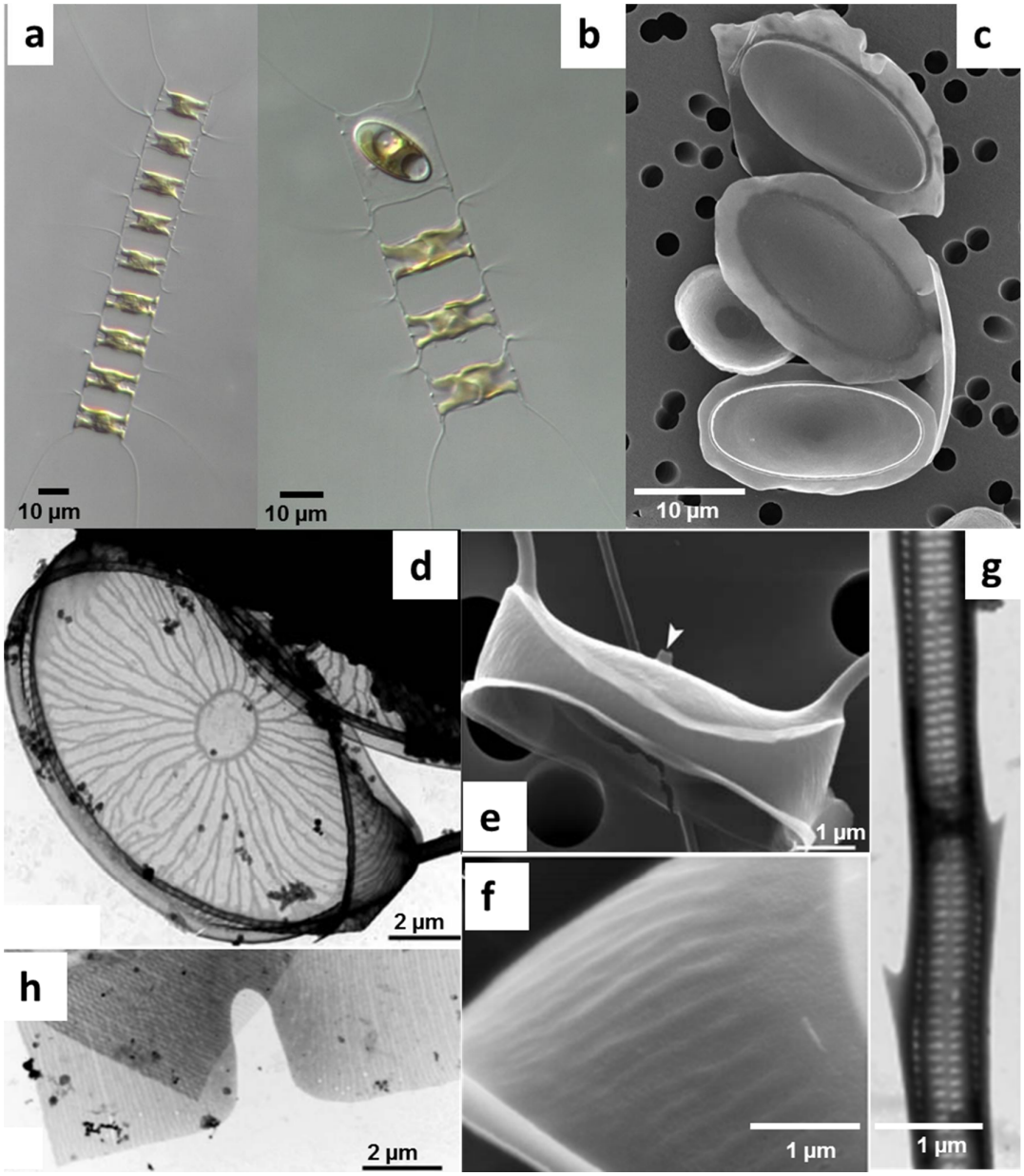


Fig. 3.37. *Chaetoceros rotosporus*. Strain Na22A3 Figs (a−c) and Figs (d−g) are reproduced from Li *et al.* (2013). LM (a−b), SEM (c and e−f) and TEM (d and g−h). (a) A colony in girdle view. (b) A colony with three vegetative cells and one resting spore. (c) Resting spores. (d) Intercalary valve. (e) Terminal valve (arrow mark tube-shaped rimoportula). (f) Parallel ribs on the mantle. (g) Detail of seta with four longitudinal rows of poroids and spines. (h) Girdle bands with parallel costae.

***C. seiracanthus* Gran**

Fig. 3.38. a−b

The described morphology corresponds to the strain newEB3big. The gross morphology of this strain matched the description provided by Jensen & Moestrup (1998). Chains short, with wide apertures between adjacent cells (Fig. 3.38. a). Each cell contains a single chloroplast. Intercalary setae emerge on the valve corner with a basal part and cross just on the chain margin and are directed in all directions (Fig. 3.38. b). Terminal setae straight, almost parallel to the axis of the colony and generally diverging from the apical plane. Resting spores vaulted in the centre and flatter at the margin, with vault on primary valve raised higher than secondary valve (Fig. 3.38. b). Vaults ornamented with long and thin spines at central part; a series of spines at the margin of primary valve. Ultrastructural details were not observed for this strain.

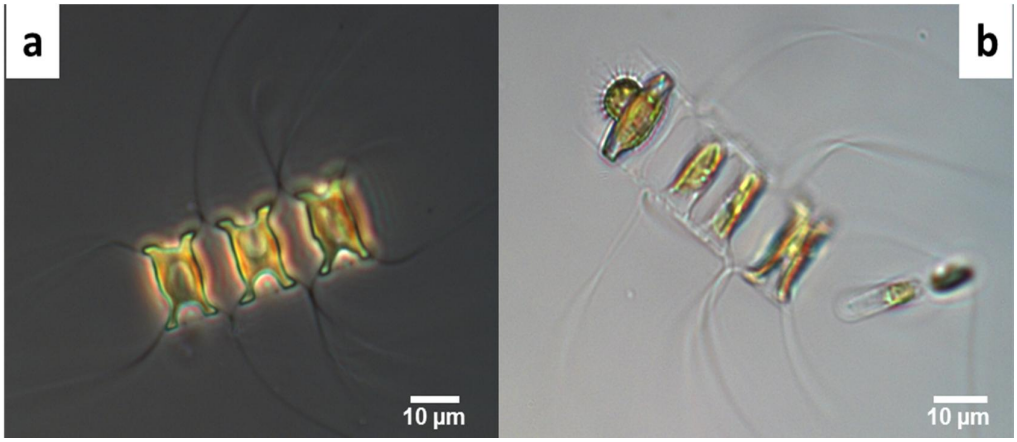


Fig. 3.38. *Chaetoceros seiracanthus*. Strain newEB3big Figs (a−b). LM (a−b). (a) A colony in girdle view (b) A colony with a resting spore.

C. 'singlecells'

Fig. 3.39. a−c

One strain was collected (Na17B2) and examined only in LM. Solitary cells are squared in girdle view and possess a single chloroplast (Figs 3.39. a−b). In culture material, cells remain in groups after cell division but the setae are not fused (Fig. 3.39. c). Setae arise from the valve corners with a short basal part and run parallel to cell axis with a U-shaped appearance.

Resting spores were not seen. Ultrastructural details were not obtained for this strain. Only 18S sequences could be obtained for this strain.

The name 'singlecells' is provisional.

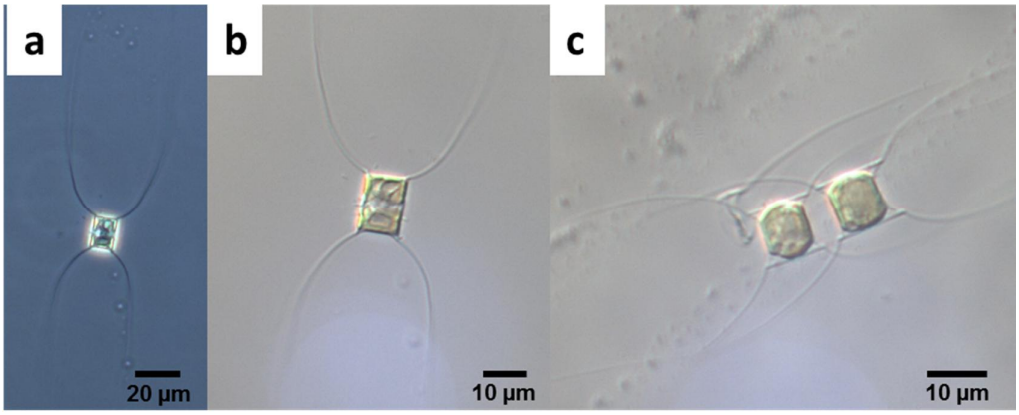


Fig. 3.39. *Chaetoceros* 'singlecells'. Strain Na17B2 Figs (a–c). LM (a–c). (a–b) Solitary cells in girdle view (c) Two cells soon after division.

***C. socialis sensu stricto* Lauder**

A detailed description of *C. socialis sensu stricto* is reported in Chapter 5.

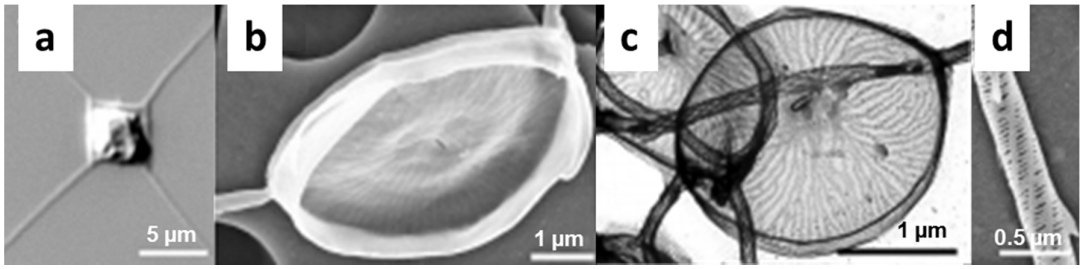


Fig. 3.40. *Chaetoceros tenuissimus*. Strain SZN-B429 Figs (a–d) are reproduced from Kooistra *et al.* (2010). LM (a), SEM (b, d) and TEM (c). (a) Solitary cell in girdle view. (b) Internal view of a valve, with the slit-shaped process. (c) Valve with slit-shaped process (d) Detail of a seta.

***C. tenuissimus* Meunier**

Fig. 3.40. a–d

The sequences and morphology of the strains collected in this study match with those of *C. tenuissimus* (strain SZN-B429) in Kooistra *et al.* (2010). The illustration provided here is taken from Kooistra *et al.* (2010). Solitary cells, appearing square in girdle view and possess single chloroplast (Fig. 3.40. a). Setae arise from the valve corners, forming a 45° angle to

apical axis and pervalvar axis. Setae circular in cross section with spirally arranged spine and pores (Fig. 3.40. d). Valves elliptical in outline, with radially branched costae and a central slit−like process, is present internally (Figs 3.40. b−c). Resting spores not observed.

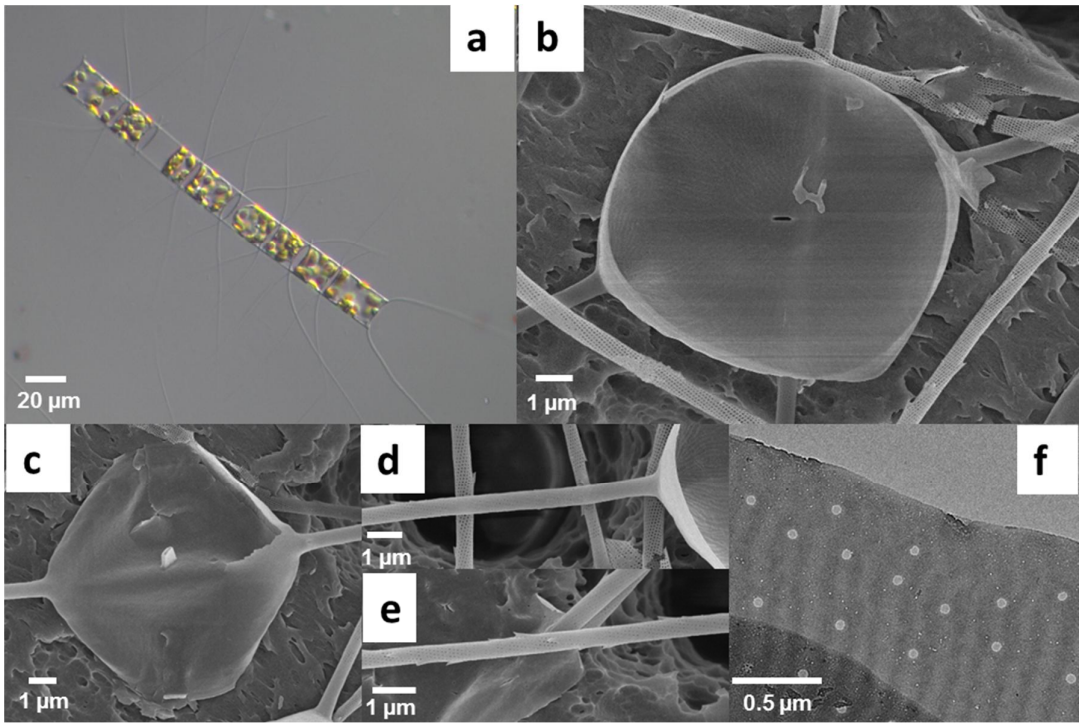


Fig. 3.41. *Chaetoceros teres*. Strain Ch2C3 Figs (b−f) and Ch12B1 Fig. (a). LM (a), SEM (b−e). TEM (f). (a) A colony in girdle view (b) Internal view of the terminal valve with the central slit-shaped rimoportula. (c) External view of the terminal valve with the external process of the rimoportula. (d) Detail of seta base (e) Detail of the middle part of a seta. (f) Girdle band with transverse costae and scattered poroids.

C. teres **Cleve**

Fig. 3.41. a−f

Eight strains were isolated in this study. Their morphology matched the description provided by Jensen & Moestrup (1998) and Shevchenko *et al.* (2006). Illustrated morphology is of strains Ch2C3 and Ch12B1. Chains usually short with 4−8 cells. Aperture varies from narrow to wide lanceolate (Fig. 3.41. a). Each cell contains several chloroplasts. Intercalary setae long, emerge on the valve corner, cross on the chain margin, and are oriented in all directions. Terminal setae U-shaped and orientated almost parallel to the axis of the colony. Setae circular in cross-section and possess numerous spirally arranged poroids and shark-fin

spines (Figs 3.41. d−e). Valves elliptical to circular in outline, ornamented with radially branched costae. Terminal valves show a central slit-shaped rimoportula (Fig. 3.41. b) with an external tubular projection (Fig. 3.41. c). Girdle bands with transverse costae and scattered poroids (Fig. 3.41. e). Resting spores not seen in the culture material.

C. tortissimus Gran

Fig. 3.42. a−c

During this study, three strains were collected. The gross morphology of these strains matched the description provided by Shevchenko *et al.* (2006). Illustrated morphology is of the strain Na25B2. Cells united to form long, strongly twisted chains (Fig. 3.42. a), in such a manner that the sibling valves appear perpendicular to one another (Fig. 3.42. b). Each cell possesses one chloroplast. Apertures narrow, hexagonal in shape. Intercalary setae originate from the valve apices and fuse immediately and cross over on the chain axis. Terminal setae extend almost parallel to the colony axis in a U-shaped (Fig. 3.42. c). Cells appear broadly elliptical in valve view and compressed in trans-apical direction, in girdle view they appear rectangular with rounded valve corners. Valve face slightly convex, mantle sometimes constricted near the margin and girdle equidimensional or higher than the mantle. Resting spores not seen in culture material. Ultrastructural details were not observed for this strain.

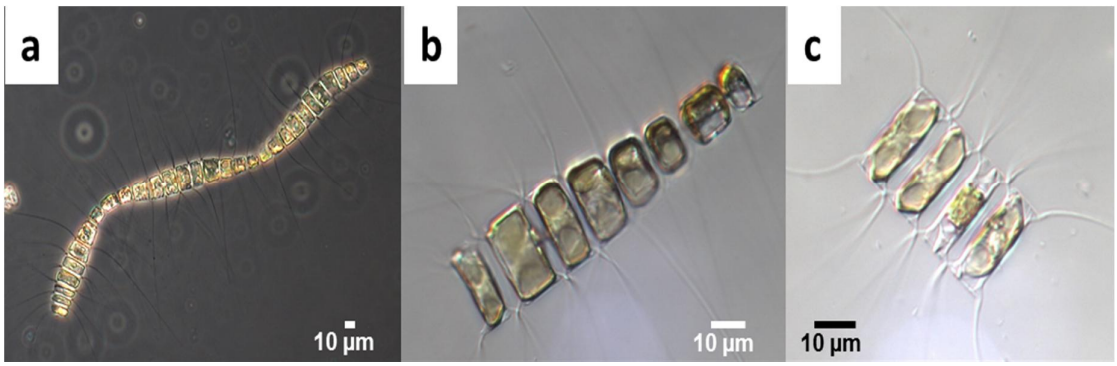


Fig. 3.42. *Chaetoceros tortissimus*. Strain Na25B2 Figs (a−c). LM (a−c). (a) A colony in girdle view. (b) Short chain with a high degree of torsion at the end of the chain. (c) Short chain without torsion.

C. 'verylongsetae'

Fig. 3.43. a–b

Only one strain (Na13C2) was isolated. Cells forming straight chains, with wide hexagonal apertures and concave valve face, with a prominent central inflation (Fig. 3.43. a). Each cell contains a single chloroplast. Mantle low or equidimensional with the girdle, a slight constriction is visible near the suture (Fig. 3.43. b). Long intercalary setae originate from the valve corners and cross slightly outside the valve margin, and oriented in all directions. Terminal setae straight, and generally diverge from the chain axis giving a U-shaped appearance. Ultrastructural details not obtained for this strain. The name 'verylongsetae' is provisional.

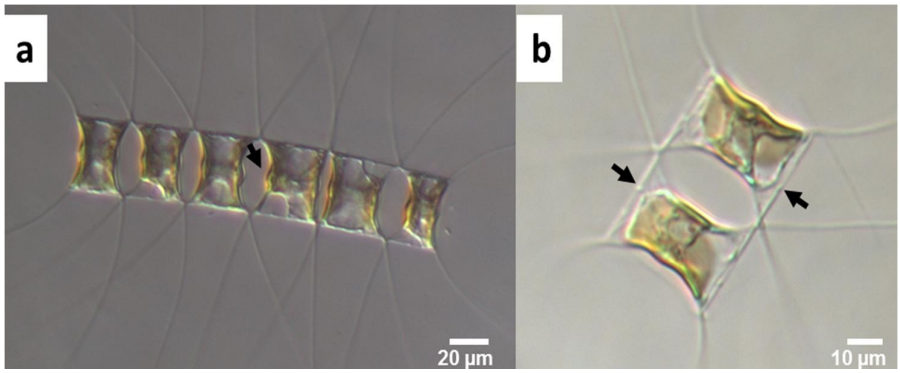


Fig. 3.43. *Chaetoceros* 'verylongsetae'. Strain Na13C2 Figs (a–b). LM (a–b). (a) A colony in girdle view (arrow mark inflation) (b) Aperture (arrow mark a slight constriction between mantle and girdle).

C. vixvisibilis Schiller

Fig. 3.44. a–i

Described morphology corresponds to the strain Na18C2 and Na25C3. Chains long, bent, with narrow lanceolate apertures (Fig. 3.44. a). Each cell contains two chloroplasts (Fig. 3.44. b). Intercalary setae long and emerge from the corners of the valves without the terminal part, fuse immediately and directed perpendicular to chain axis (Fig. 3.44. g). Terminal setae perpendicular to chain axis. Setae circular in cross-section and possess

spirally organised shark-fin spines and rows of poroids, with large pores distributed regularly (Figs 3.44. h–i). Valve elliptical in outline, ornamented with dichotomously branched costae. Terminal valve possesses with a central rimoportula externally (Figs 3.44. e–f), and a small slit internally (Fig. 3.44. d). A hyaline rim is present on the marginal ridge (Fig. 3.44. c). Resting spores not observed.

The Neapolitan strains showed minor differences from that reported by Hernández-Becerril *et al.* (2010): the ultrastructure of the setae and the chloroplast number. Hernández-Becerril *et al.* (2010) also described the resting spores, which are convex, dome-shaped. Their primary valve is reported as smooth and possesses long spine/s (1-4), sometimes branched, and the secondary valve is smooth without any structures.

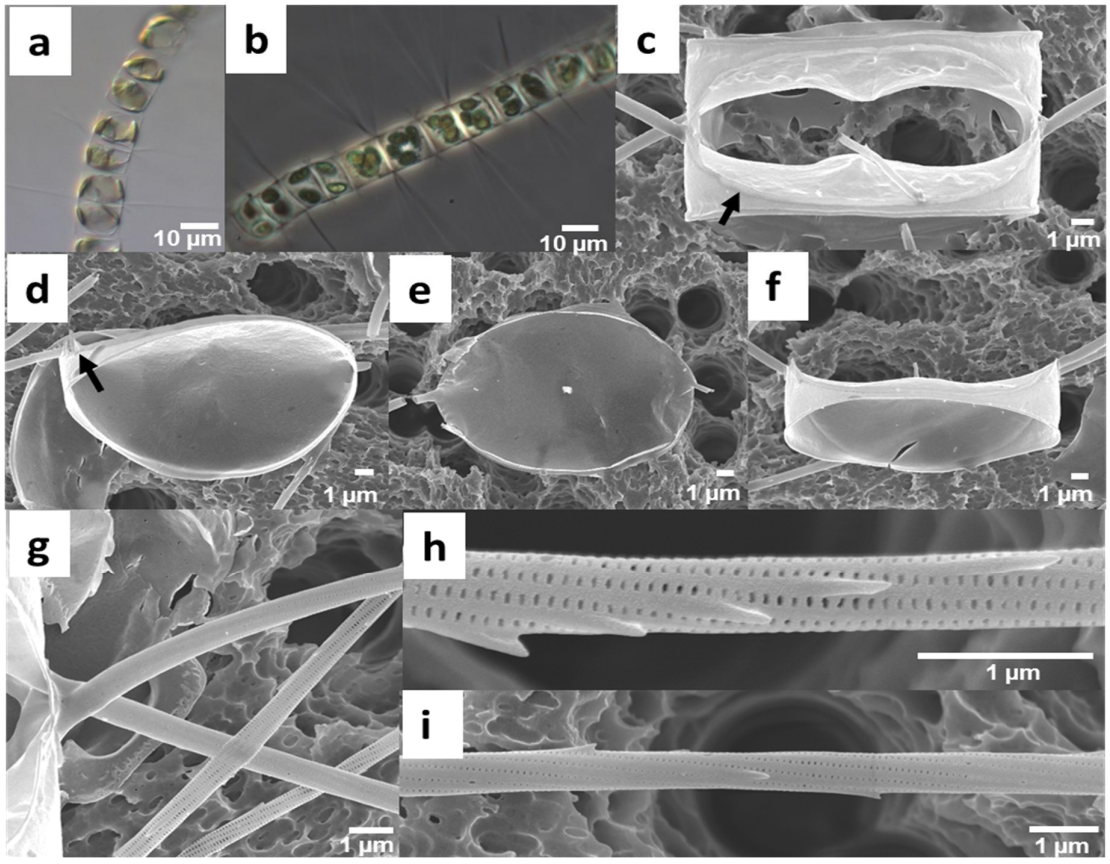


Fig. 3.44. *Chaetoceros vixvisibilis*. Strains Na18C2 Figs (c–i) and Na25C3 Figs (a–b). LM (a–b), SEM (c–i). (a–b) A colony in girdle view. (c) Aperture (arrow mark small hyaline marginal rim). (d) Internal view of a terminal valve with a central slit-shaped process (arrow indicate horn on the base of terminal setae). (e–f) External view of terminal valves with the tubular process of the rimoportula. (g) Detailed structure of setae base note the base of setae is smooth. (h–i) Details of setae.

b) Phylogenetic study

PCR primers: The pairs of universal 28S and 18S PCR primers generated ca.750 bp and ca.1800 bp DNA PCR products, respectively (see Chapter 2). In some cases, the product was markedly longer due to the presence of inserts (see Chapter 6 for details on these inserts). Few strains failed to amplify 18S with the universal primers. In such cases, internal primers were used to obtain a set of shorter, but overlapping, DNA templates of various sizes, sequenced individually and then aligned to generate the entire sequence. Nonetheless, I failed to obtain the marker sequences of the cultured species *C. affinis* and *C. constrictus*.

Table 3.2. The V4-primer sites in 18S rDNA and its misfits. The primer positions are depicted as present in the alignment in the forward reading frame. The blue nucleotides indicate ambiguity and the red ones represent substitutions.

Primer	CCAGCASCYGCGGTAATTCC	TCATYRATCAAGAACGAAAGT
C. muellerii	CCAGCAGCCGCGGTAATTCC	TCATTGGTCAAGAACGAAAGT
C. anastomosans	CCAGCAGCCGCGGTAATTCC	TCATTGGTCAAGAACGAAAGT
C. vixvisibilis	CCAGCAGCCGCGGTAATTCC	TCATTGGTCAAGAACGAAAGT
C. cf. *affinis*-V	CCAGCAGCCGCGGTAATTCC	TCATTGGTCAAGAACGAAAGT
C. cf. *dayaensis*	CCAGCAGCCGCGGTAATTCC	TCATTGGTCAAGAACGAAAGT
C. radicans	CCAGCAGCCGCGGTAATTCC	TGTTTGGTCAAGAACTAAAGT
C. cinctus	CCAGCAGCCGCGGTAATTCC	TAATTGATCAAGAACGAAAGT

The V4-forward primer did fit the sequence in all Chaetocerotaceae and outgroups, whereas the reverse primers had few misfits that may possibly hinder the DNA amplification in some species of *C. anastomosans, C. vixvisibilis, C.* cf. *dayaensis, C.* cf. *affinis, C. radicans* and *C. cinctus* (Table 3.2). The most ruptured primer positions were observed for *C. radicans* with several misfits towards the 3'-end, which can be deleterious for PCR amplification.

Sequence alignment: Alignment of the 105 selected 18S sequences (77 ingroup sequences) resulted in a dataset of 1788 base positions, including the 4 positions (N) for each one of nine short inserts and 8 positions (N) for each one of three long inserts (see Chapter 6). Of these positions, 845 were constant (including the sixty (9x4+3x8) characters indicated as 'N'), 182 were variable but parsimony uninformative and 761 were parsimony informative.

Alignment of the 94 selected partial 28S sequences resulted in a dataset of 768 base positions, with a presence of an insert of ca. 200 bp at position 735 in the alignment. Excluding the insert, 289 positions were constant, 45 were variable but parsimony uninformative, and 434 were parsimony informative.

The V4 region of the Chaetocerotacean reference- and out-group sequences ranged between 373bp (*C. mullerii*) and 383bp (*C. curvisetus*). Alignment of the 105 selected V4 sequences resulted in a dataset of 395 base positions. The sequences of *C.* 'verylongsetae' Na13C2 and *C. diversus* Na5B2 exhibited an insert of 130 and 120bp at positions 323 and 325 respectively, rendering their length to 510 and 500bp respectively. *B. hyalinum* strain CCMP141 showed a presence of insert at position 323 while the strains obtained from the study are lacked an insert. Excluding this insert, 158 positions were constant, 30 were variable but parsimony uninformative and 205 parsimony informative.

Phylogeny inferred from the 18S rDNA sequences: trees were generated using maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI).

For the inference of the ML tree, the chosen model was GTR with a base substitution rate matrix of A↔C (1.0241), A↔G (2.7501), A↔T (1.2891), C↔G (0.7110), C↔T (4.6968), relative to G↔T (1.0000); gamma-distributed rate variation among sites with shape parameter $\alpha = 0.3258$ and estimated base frequencies A 0.270, C 0.191, G 0.256, and T 0.283. The likelihood score (-ln L) for the ML tree for 18S alignment was 25388.9572. Bayesian inference was conducted with 1 million MCMC runs. The posterior probability of the sampled trees showed a standard deviation value of 0.010. After discarding 25% of the initially sampled trees as burn-in, and 7501 trees were sampled for my inference. Six equally MP trees were retained, of length 5466, a rescaled consistency index 0.306, a homoplasy index 0.677 and a G-fit of -431.521. The only difference among these six trees was in the position of the three sequences in the *C. peruvianus-rostratus* clade. Only the ML tree was

considered for the inference with the bootstrap valves (ML/MP) and bpp (BI) depicted on the resulting tree ML tree (Fig. 3.45).

The resulting 18S trees all resolved a Chaetocerotacean clade (see Figs 3.45) with weak bootstrap support (68% in ML, 53% in MP) and high posterior probability (100%) in BI. The outgroup sequences of *Dactyliosolen* were recovered in a clade as sister to Chaetocerotaceae, likewise with weak bootstrap support (66% in ML, 67% in MP) and high posterior probability (100%) in BI. The Chaetocerotacean clade showed a poorly resolved basal topology from which five well-supported clades emerged.

Clade I: in the ML-tree, the first of these Chaetocerotacean clades (99%) (from the top downwards) contained in its turn a clade (97%) with sequences of the members of subgenus *Chaetoceros* as sister to a clade (100%) with *C. didymus* and *C. protuberans* (belonging to subgenus *Hyalochaete*). *Chaetoceros eibenii* was the first to branch off from the remainder of the subgenus *Chaetoceros* with high bootstrap support.

Clade II: the second clade (100%) included all sequences of the *Bacteriastrum* species, with the one of *B. hyalinum* as sister to a well-supported clade (100%) with the sequences of the five other species included in the analysis.

Clade III: the third clade (100%) included sequences belonging to strains of *C. contortus*.

Clade IV: the fourth clade (100%) consisted in its turn of several clades, including from top to bottom; a lineage with *C.* cf. *constrictus*, a clade (100%) with sequences of *C. diadema*, *C. rotosporus*, *C. seiracantus* and *C.* cf. *holsaticus*, a clade (71%) with sequences of *C. muelleri*, *C. decipiens* and several sequences belonging to the *C. lorenzianus* complex, a clade (100%) with sequences of *C. anastomosans, C.* cf. *dayaensis, C. vixvisibilis* and several species new to science, and a clade (99%) with sequences of *C. diversus*, *C. circinalis, C.* cf. *pseudocrinitus,* and two unidentified *Chaetoceros* strains. Topology among these clades was only weakly supported or not at all.

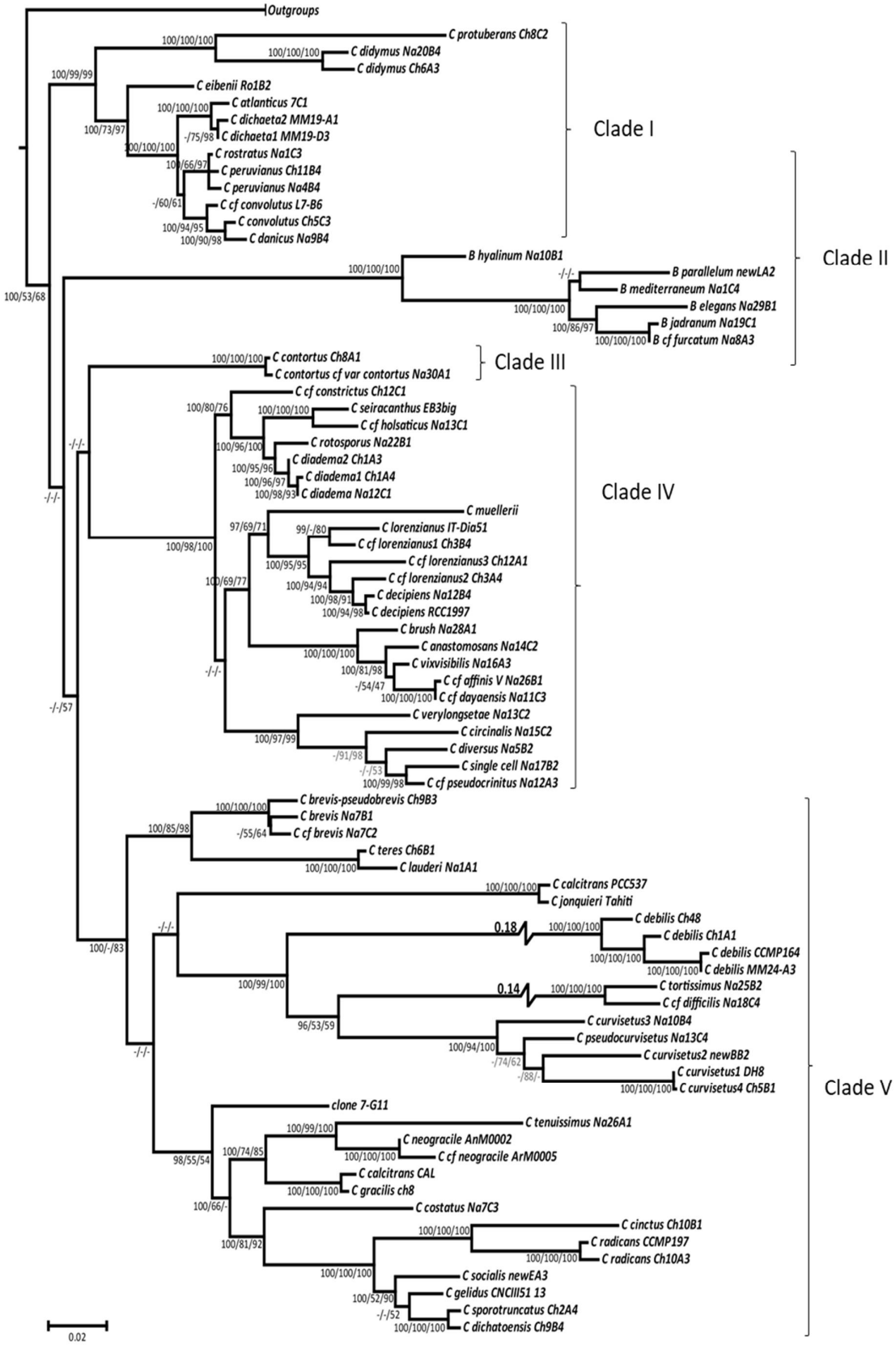


Fig. 3.45. Maximum likelihood tree inferred from 18S rDNA sequences of Chaetocerotacean strains. Far outgroups have been pruned away. Bootstrap values are indicated (ML/MP/BI) at their respective internodes. Bootstrap values below 50% have been marked as "-".

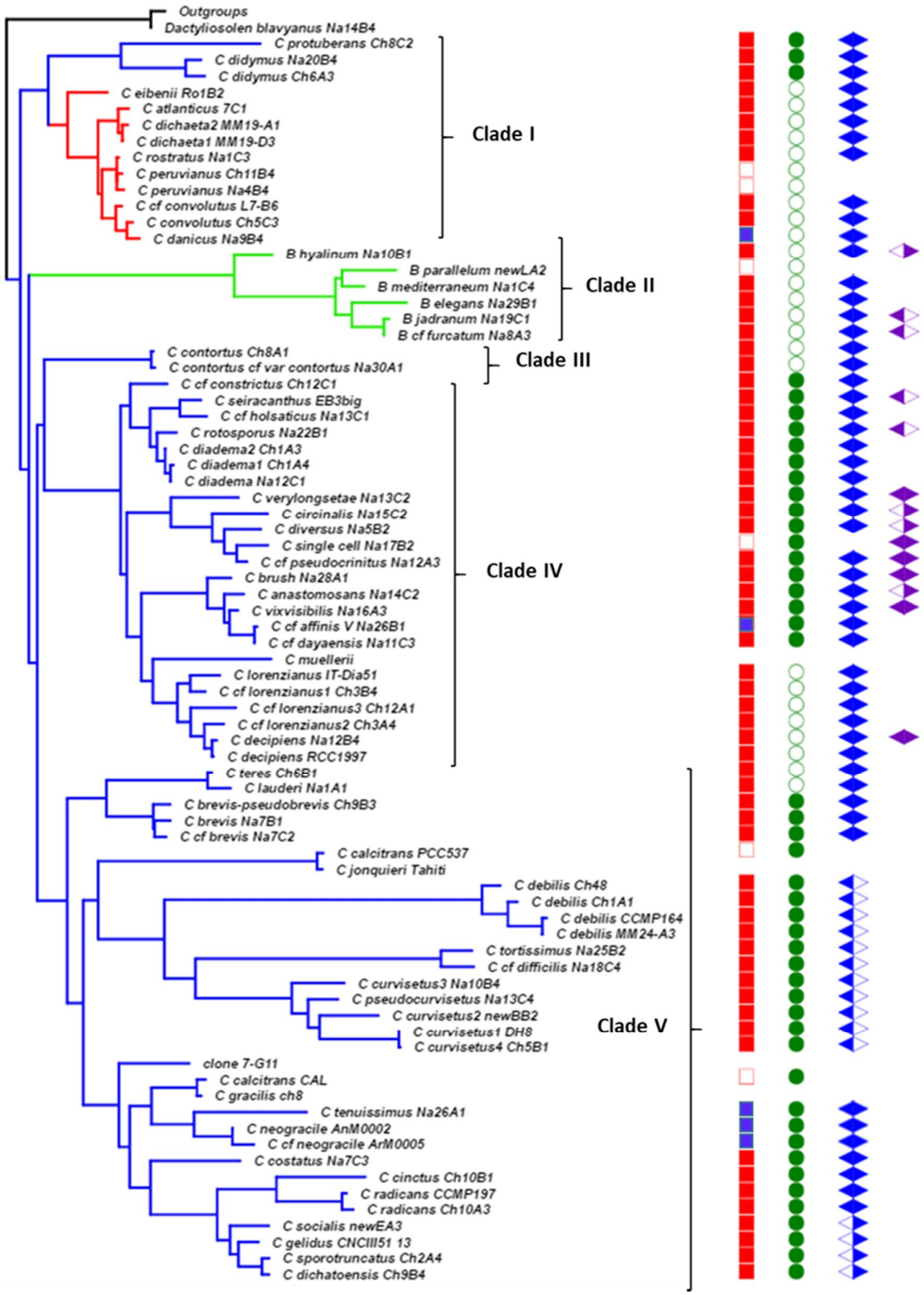


Fig. 3.45.1. Colored clades and grades indicate groups: The blue grade represents subgenus *Hyalochaete*; the red clade, subgenus *Chaetoceros*, and the green clade, genus *Bacteriastrum*. Squares indicate colony morphology: empty squares, single cells; red squares, colonies; blue squares indicate "both morphologies observed." Circles indicate chloroplast numbers: empty green circles, single plastids; filled-in green circles, multiple plastids. Colony shape is represent as 'straight or bent' (blue diamond), vs 'twisted or tortion' (leftside filled blue diamond), 'curved or helical' (right side filled blue diamond). The purple diamond indicates presence of short and long introns, rightside filled blue diamond for short and leftside filled for long introns. Blank space indicates no information available or no such characteristics.

Clade V: The fifth clade (83%) included in its turn several major clades in a badly resolved basal topology. The first of these clades (98%) included sequences of *C. brevis*/*pseudobrevis* like species as well as *C. teres* and *C. lauderi*. The second clade (100%) contained sequences of *C. calcitrans* and *C. jonquieri*. The third clade (100%) contained a clade with sequences of *C. debilis* as sister to a clade with sequences of *C. tortissimus* and *C.* cf. *difficilis* and a clade with sequences of *C. pseudocurvisetus* and *C. curvisetus*. The fourth clade included an environmental clone 7-G11 and a clade with the single-celled species *C. tenuissimus, C. neogracile, C. calcitrans* and *C. gracilis* as sister to a clade with the sequences of *C. costatus*, *C. radicans*. *C. cinctus* and the *C. socialis* complex.

Information on morphological characteristics, colony formation, number of chloroplast, chain morphology, presence/absence of introns in species and the subgenus to which taxa belong to are represented in Fig. 3.45.1.

The topologies among the five clades in the MP and BI trees corroborated that as recovered in the ML tree with the following exception. In the MP tree the fifth clade as recovered in the ML-tree lacked sufficient support, and also its four daughter clades collapsed basically onto the Chaetocerotacean basal polytomy.

Phylogeny inferred from the partial 28S rDNA sequences: Phylogenetic trees were generated using maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI). Regarding the ML tree, the chosen model was GTR with a base substitution rates matrix of A↔C (0.8623), A↔G (2.6094), A↔T (1.2404), C↔G (0.6330), C↔T (5.1834), relative to G↔T (set at 1.0000); the gamma-distributed rate variation among sites with $\alpha = 0.4509$ and estimated base frequencies A 0.267, C 0.188, G 0.290, and T 0.256. The likelihood score (-ln L) for the ML tree for partial 28S alignment was 14348.8786. The best tree is shown in Fig. 3.46. The Bayesian inference was terminated after 1 million MCMC runs. The posterior probability of the sampled trees showed a standard deviation value of

0.011. Approximately, 25% of the initially sampled trees were discarded as burn-in, and 7501 trees were sampled for my inference. The MP-analysis resulted in two equally MP trees with length 2885, a rescaled consistency index of 0.234, a homoplasy index of 0.693 and a G-fit of -235.6505. The only difference among the trees was in the position of the sequences in the clade of *C. compressus-contortus*. Only the ML tree is considered for the inference with the bootstrap valves (ML/MP) and bpp (BI) depicted on the resulting tree ML tree (Fig. 3.46).

All analyses recovered the major five Chaetocerotacean clades as observed in the 18S rDNA trees and, likewise, lacked support for any phylogenetic structure among them. However, in contrast to the state of affairs in the 18S trees, the outgroup sequences did not form a reciprocal sister clade to Chaetocerotaceae.

In the ML-tree, *Dactyliosolen blavyanus* and *Hemiaulus hauckii* were recovered as sister taxa, but a clade (100%) with *Eucampia* spp was recovered away from it, as sister to a large clade of *Chaetoceros* species, though only with marginal support (50%), thus rendering Chaetocerotaceae marginally paraphyletic. The basal topology (or lack thereof) was encountered in the MP and BI trees.

Clade I: in the ML-tree, the first Chaetocerotacean clade (98%) included sequences belonging to strains of *C. contortus*/*compressus*. The clade contained six sequences retrieved from different geographical areas. In Gulf of Naples, Italy (GoN), two varieties of *C. contortus* were found.

Clade II: the second Chaetocerotacean clade (100%) contained a well-supported clade (100%) with *C. didymus* and *C. protuberans* (subgenus *Hyalochaete*) and a clade (100%) with sequences of subgenus *Chaetoceros,* except *C. eibenii.* Grouping of the latter with the other sequences of subgenus *Chaetoceros* did not obtain sufficient bootstrap support.

Clade III: the third Chaetocerotacean clade (83%) consisted of a polytomy of several clades including a lineage with *C*. cf. *constrictus*, a clade with sequences of *C. affinis*, *C. diversus*, *C. circinalis, C*. cf. *pseudocrinitus, C*. cf. *simplex* and two unidentified *Chaetoceros* strains, a clade with sequences of *C. diadema, C. rotosporus, C. seiracantus* and *C.* cf. *holsaticus*, a clade with sequences of *C*. cf. *minimus, C*. cf. *laevisporus*, *C*. *decipiens* and several sequences belonging to the *C*. *lorenzianus* complex, a clade with sequences of *C. anastomosans*, *C. dayaensis*, *C. vixvisibilis* and several still undescribed species.

Clade IV: the fourth clade (100%) included the sequences of the seven *Bacteriastrum* species included in this study. The sequence of *B.* cf. *furcatum* strain Na8A3 from the Tyrrhenian Sea differed from that of strain *B. furcatum* PMF-BA4 from the Adriatic Sea (Bosak *et al.* 2015); the former grouping as sister to *B. jadranum* and the latter close to *B. hyalinum*.

Clade V: the fifth clade (99%) included in turn three major clades. The first of which included sequences of *C. laciniosus*, *C. brevis*/*pseudobrevis* like species, *C*. *teres* and *C*. *lauderi*. The second of these clades contained sequences of *C. debilis*, *C. tortissimus*, *C*. cf. *difficilis*, *C. pseudocurvisetus* and several distinct sequences of *C*. *curvisetus*. The third of these clades included a clade with sequences of single-celled species *C. tenuissimus, C. fallax, C. calcitrans* and *C. neogracile* as sister to a clade with sequences of *C. costatus*, *C. radicans*. *C. cinctus* and the *C. socialis* complex.

The topology of the BI tree was basically in accordance with that of the ML tree. The MP tree also showed five major Chaetocerotacean clades, but the basal topology connecting them differed from that in the ML tree in that the fourth clade (*Bacteriastrum*) resolved as sister to the fifth clade with marginal support. Within the second clade, *C. eibenii* grouped with *C. protuberans* and *C. didymus* (though, with marginal support).

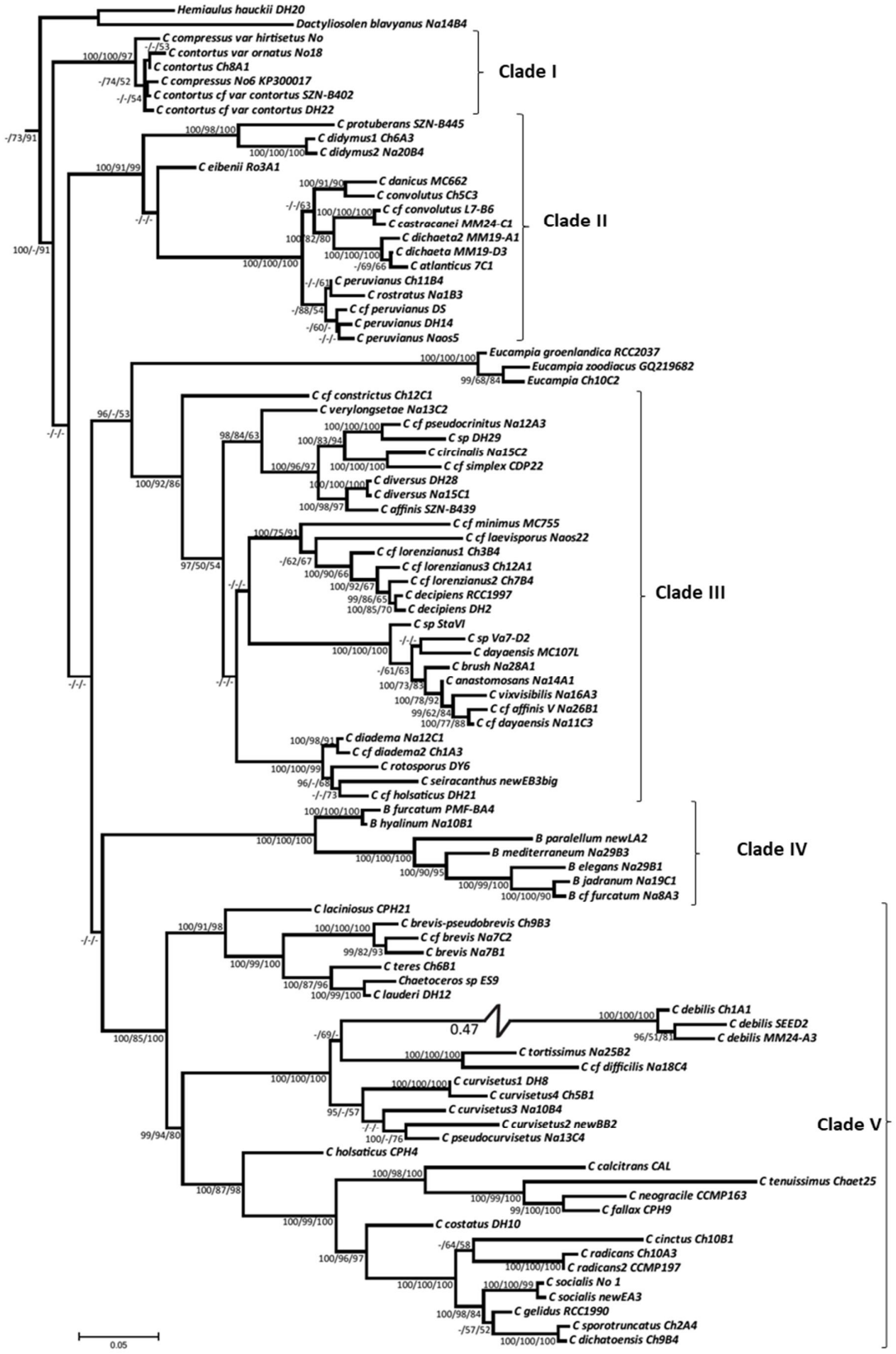


Fig. 3.46. Maximum likelihood tree inferred from the partial 28S rDNA gene region of Chaetocerotacean strains Bootstrap are indicated (ML/MP/BI) at their respective internodes. Bootstrap values below 50% have been marked as "-".

Phylogeny inferred from the V4 sequences: ML analysis done using two different softwares i.e., raxmlGUI v.1.5beta (Silvestro & Michalak 2012) and FastTree (Price et al. 2009, 2010) resulted in similar tree topology with a slightly different bootstrap values. Both trees (Fig. 3.47.1 and Fig. 3.47.2) are presented to show similar tree topologies, as FastTree will be used in the HTS analysis in next chapter. The results indicate that FastTree can be used to generate phylogenies with large dataset that may result in similar topology as RAxML, with ease of time.

Nine equally most likely trees were generated using FastTree. For the inference of the ML tree, the chosen model was GTR with a base substitution rate matrix of A↔C (0.9011), A↔G (2.5902), A↔T (0.8326), C↔G (1.1820), C↔T (4.3749), relative to G↔T (1.0000) and estimated base frequencies A 0.2644, C 0.1579, G 0.2622, and T 0.3156. The ML-Nonetheless, Chaetocerotaceae formed a clade with 88% bootstrap support. Since the clade structure of V4 trees are similar (Fig. 3.47.1 and Fig. 3.47.2), the FastTree generated ML tree was referred to describe the clades.

Clade I: within the Chaetocerotacean clade, Clade 1 (98%) included a well-supported clade with the members belonging to the subgenus *Chaetoceros* as sister to a clade with *C. didymus* and *C. protuberans* (which belong to subgenus *Hyalochaete*). The clade with subgenus *Chaetoceros* showed short internodes, with a good support for some internal ramifications, but only fair support for others.

Clade II: the second clade (99%) included only strains of *C. contortus*.

Clade III: the third clade (86%), comprised a clade of all the *Bacteriastrum* species (99%) as sister to a clade including *C. debilis* (81%), *C. tortissimus*, *C.* cf. *difficilis, C. pseudocurvisetus* and three cryptic species of *C. curvisetus*.

Clade IV: the fourth clade (99%) contained in its turn, a polytomy from which several clades emerged. The first of these included *C.* cf. *constrictus*, the second contained *C. circinalis*, *C.*

cf. *pseudocrinitus*, *C. diversus* and two unidentified *Chaetoceros* strains (97%), the third (94%) included *C. seiracantus*, *C*. cf. *holsaticus, C. diadema* and *C. rotosporus*, and the last clade (94%) consisted in its turn of a clade with sequences of *C. anastomosans*, *C. vixvisibilis*, and several Neapolitan strains of species new to science as sister to a clade with *C. decipiens* and several species belonging to the *C. lorenzianus* complex.

Clade V: the basal topology of the fifth clade was poorly supported. This clade contained in its turn several well-supported clades. One of these clades (91%) included in its turn two highly supported clades, one with *C. lauderi* and *C. teres* and the other one with species within the *C. brevis* - *C. pseudobrevis* complex. One clade (100%) included sequences of *C. jonquieri* and *C. calcitrans*. One included only an environmental clone, 7-G11. In addition, one clade (86%) contained in its turn a clade with sequences of *C. calcitrans, C. gracilis, C. neogracile* and *C. tenuissimus* as sister to a clade containing *C. costatus*, *C. radicans, C. cinctus, C. gelidus* and the *C. socialis* complex.

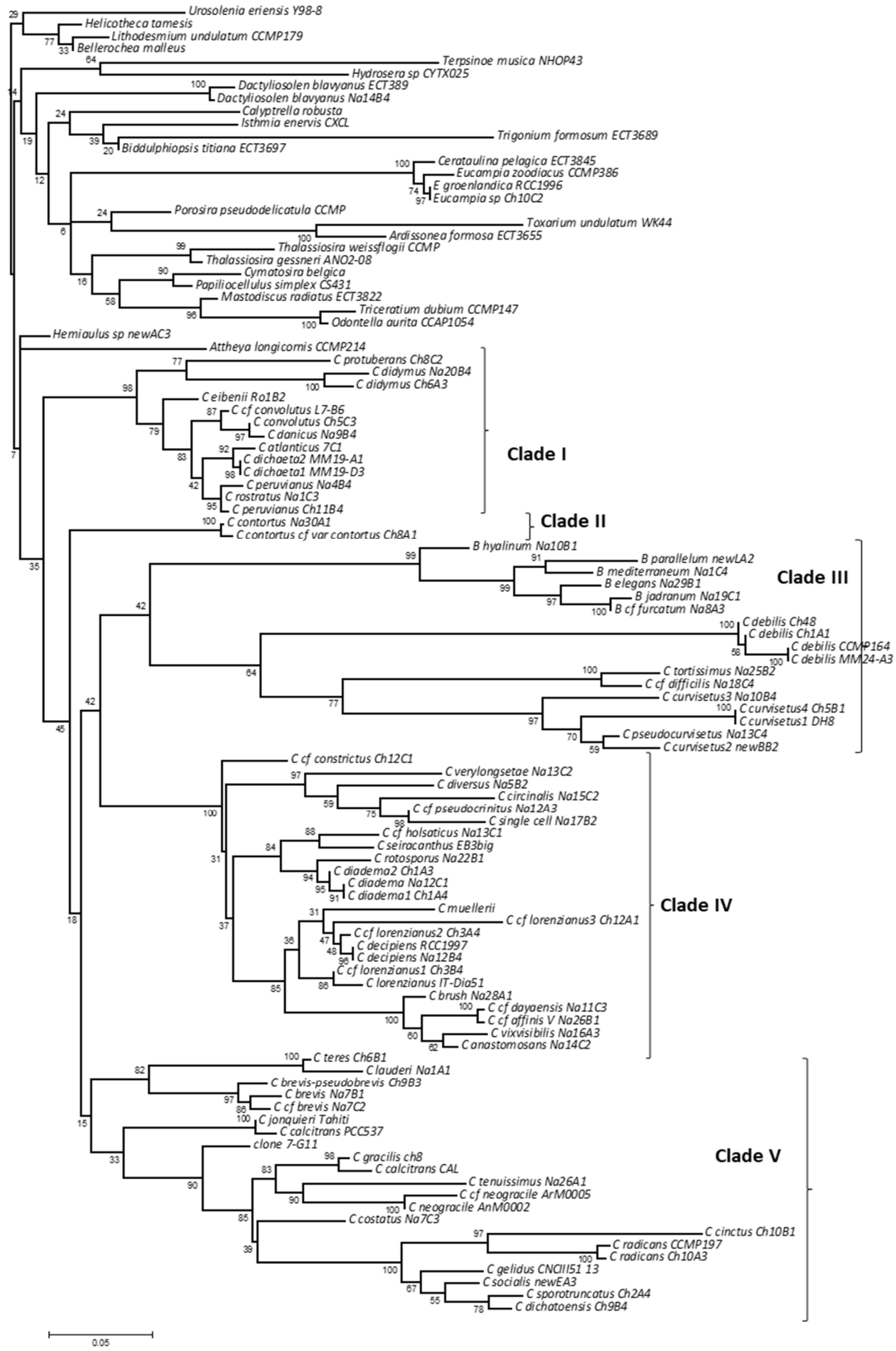


Fig. 3.47.1. Maximum likelihood tree inferred from the V4 region of the 18S rDNA sequences of Chaetocerotacean strains using RaxML. Bootstrap support values are represented as percentages.

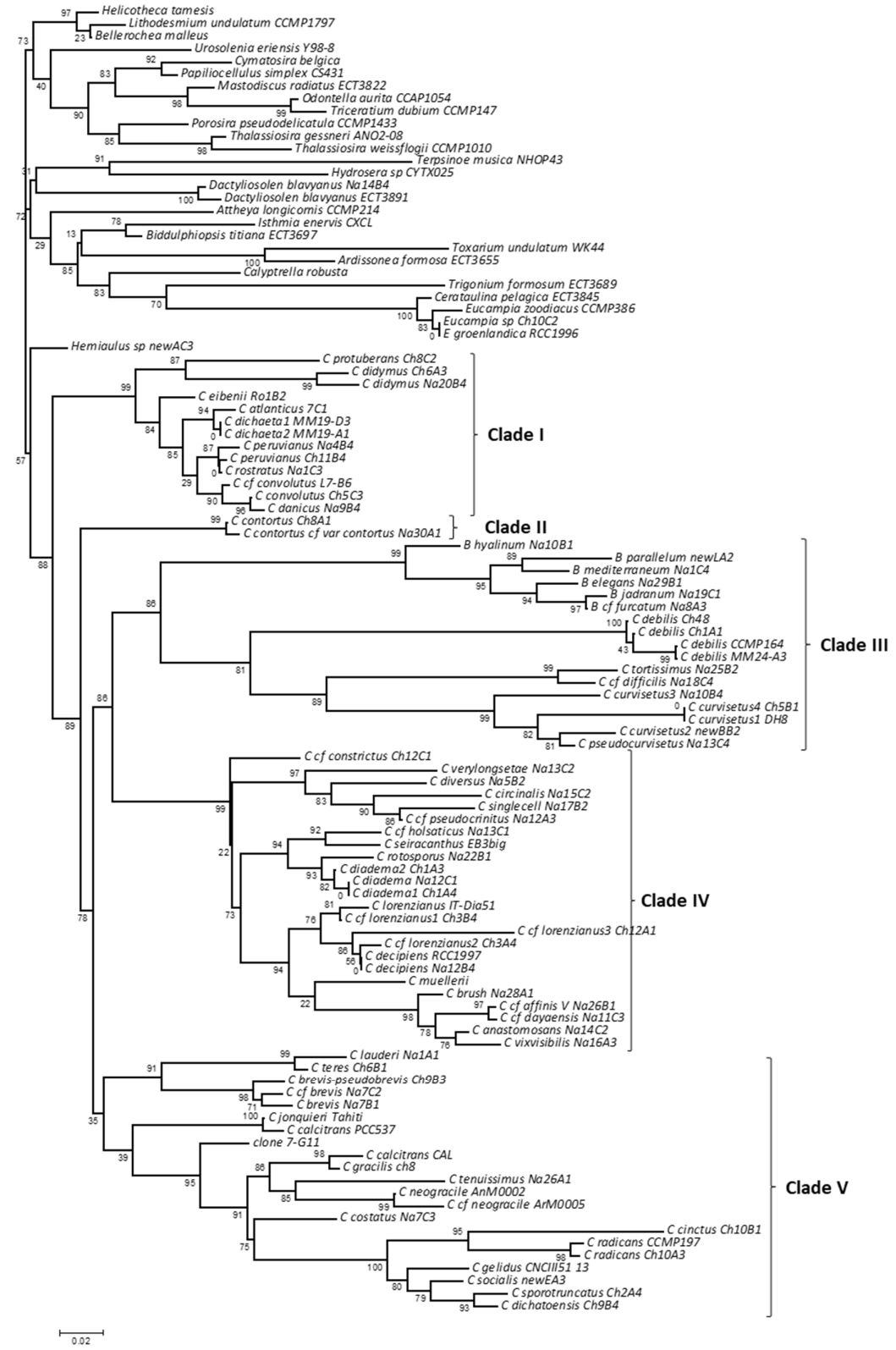


Fig. 3.47.2. Maximum likelihood tree of the V4 region of the 18S rDNA sequences of Chaetocerotacean strains using FastTree. Bootstrap support values are represented as percentages. Note, the topology of the trees generated using FastTree are similar, with a few changes in Bootstrap values. This provides information that FastTree can be used for inferring phylogeny for bigger datasets.

Discussion

The results of the present study reveal that even at a taxonomically well-studied plankton monitoring station such as the LTER MareChiara (LTER-MC) numerous species new to science can still be discovered. Several of these belong to complexes of morphologically similar or even cryptic species, difficult if not impossible to tell apart in routine microscopical surveys. Several species not known from the LTER-MC, described recently from elsewhere (e.g., *C. rotosporus*) are surprisingly common at the LTER-MC. Yet, they can be detected by applying culture-based or genetic approaches. The occurrence of all these similar species shows that morphological traits previously assumed to be defining single species are actually shared among several species. Therefore, additional traits are needed to identify these species morphologically.

The results confirm that at the LTER-MC, the genus *Chaetoceros* is indeed, far more diverse than *Bacteriastrum*, that the latter is monophyletic inside paraphyletic *Chaetoceros*, and that subgenus *Chaetoceros* forms a clade inside a paraphyletic subgenus *Hyalochaete*.

Another major outcome is that most of the Chaetocerotacean diversity can be detected also applying a high-throughput sequencing metabarcoding approach using the V4 region in the 18S rDNA as barcode marker. Almost all of the species detected in the present study can be discriminated using this marker. Inserts appear in multiple positions along the 18S and 28S rDNA, but only two species exhibit a ca. 110 bp. insert in their V4 barcode region, meaning that high-throughput sequencing will detect almost all of the species. Last but not the least, the universal primers used to amplify the V4 barcode fit the primer targets of all of the species, except a few that never have been observed in the plankton counts at the LTER-MC.

Phylogeny of Chaetocerotaceae

The phylogenies inferred from partial 28S and 18S rDNA sequences basically reveal a similar topology with the major clades. Regardless of the sequences or the phylogenetic

algorithms (MP, ML, and BI), the resulting phylogenies revealed that Chaetocerotaceae is probably monophyletic and separates into five distinct clades. The four clades include *Chaetoceros* species and the fifth one monophyletic *Bacteriastrum*. These topologies are in agreement with previously published phylogenies (Kooistra *et al.* 2010; Li *et al.* 2013, 2015; Bosak *et al.* 2015).

The 18S rDNA phylogeny includes only relatively near outgroups in the polar centric diatoms to mitigate possible long-branch attraction (LBA) artifacts (Theriot *et al.* 2009). Sequences of close outgroup taxa *Ceratulina*, *Dactyliosolen*, *Eucampia*, *Hemiaulus*, *Hydrosera*, *Terpinoe*, *Urosolenia* as well as more distant outgroup taxa (*Attheya*, *Biddulphiopsis*, *Helicotheca*, *Lithodesmium*, *Odontella* and *Thalassiosira*) known to form relatively short branches in phylogenetic trees, are included (Ashworth *et al.* 2013; and personal findings). The phylogeny inferred from partial 28S rDNA sequences includes only the closest among these outgroups as recovered in the 18S tree. One reason is that 28S sequences are unavailable for many of the above-mentioned genera, and moreover, more distant outgroups reduce the resolution of the tree topology and support for internodes.

Most of the major clades show a well-resolved internal structure but a poorly resolved basal structure. The most basal branching pattern below the five major clades did receive poor support if any at all. This might have resulted from adaptive radiation. Once the last common ancestor of the family acquired setae, it may have diversified into several different ecological niches. Setae are believed to be advantageous in terms of grazer defence, buoyancy and increased surface area for nutrient uptake. Ligowski *et al.* (2012) found that *C. dichaeta* adapted to Antarctic ice by changing shape and orientation of the setae and the aperture length between sibling cells. The authors viewed this as a competition for space as a defence mechanism against predation in such habitats. Subsequent acquisition of secondary functionalities of the setae, for instance as a conduit for small chloroplasts in all members of

the monophyletic subgenus *Chaetoceros* (Kooistra *et al.* 2010), probably lead to its own rapid speciation.

Other than setae, most Chaetocerotaceae members are capable of producing resting spores that help to avoid unfavourable conditions. Thanks to the possibility to form resting spores, populations can simply hibernate through adverse periods. However, spores do not only give Chaetocerotaceae a cutting edge because these resting stages are also encountered in several other bi- and multipolar diatoms as well as in many other phytoplankton lineages.

Recovery of *Bacteriastrum* inside *Chaetoceros* is in line with the results of other studies (Rines & Theriot 2003; Kooistra *et al.* 2010; Bosak *et al.* 2015). *Bacteriastrum* inside *Chaetoceros* does not obtain high bootstrap support, though, meaning that the former is either within paraphyletic *Chaetoceros* or nearest sister to it. *Eucampia*, which groups marginally inside Chaetocerotaceae in the 28S phylogeny, groups firmly outside the family in the 18S tree. The latter is more likely because there is no evidence for setae in *Eucampia*, though the absence of setae in the latter could result from a secondary loss in this genus once the horns connecting its sister valves were acquired.

Species delineation: one of the biggest challenges in assessing the diversity in an ecosystem is a delineation of species, especially in diatoms. The present study also reveals high species richness in the family Chaetocerotaceae. Some of these species included morphologically identical species that consist of genetically distinct groups, which are reproductively isolated. Such morphologically indistinguishable, biologically distinct species are known as cryptic species (Montresor *et al.* 2003; Behnke *et al.* 2004; Sarno *et al.* 2005, 2007; Lundholm *et al.* 2006; Amato *et al.* 2007; Beszteri *et al.* 2007; Evans *et al.* 2008; Kooistra *et al.* 2008; Vanormelingen *et al.* 2008), and pseudo-cryptic, i.e., in case they can be morphologically differentiated once the genetics are known (Kooistra *et al.* 2010; Chamnansinp *et al.* 2013, 2015; Li *et al.* 2013, 2015). The work presented here shows how

combined morphological information along with sequence data, can delineate genetically and morphologically distinct entities and resolve taxonomic uncertainties.

In follow-up of the discovery of cryptic species in *Chaetoceros* by Kooistra *et al.* (2010); evidence for several cryptic species is now demonstrated among the members of *C. contortus-compressus*; *C. debilis*; *C. didymus*; *C. diadema*; *C. curvisetus*; *C. brevis*; and *C. socialis*, which occur in allopatry i.e., genetically distinct populations of the species with mere geographic variation. Also, evidence for cryptic or pseudo-cryptic diversity occurring in sympatry is confirmed in members of *C. contortus*; *C. peruvianus*; *C. affinis*; *C diversus*; *C. diadema*; *C. lorenzianus*; *C. brevis*; *C. curvisetus*; and *C. socialis*. Integrating the morphology of the isolated strain with sequence information results in delimiting a species and generating a reference barcode. However, assigning sequence data to taxonomically described taxa from the literature can be misleading if the strain from which the sequence has been obtained belongs to a different semi-cryptic species than the strain from which the morphological description has been generated.

The recognition of genetically and ultrastructurally distinct entities (*sensu stricto*) in what formerly was perceived as a single species (*sensu lato*) poses nomenclatural problems. Many original species descriptions are based solely on LM observations dating back more than a century ago and are often illustrated sketchily. To which one of the entities *sensu stricto* does the drawing refer? In addition, the original material is often no longer available for electron microscopic examination of frustule ultrastructural details. Therefore, there is no way to determine which one of the species *sensu stricto* belongs to the original "type specimen". A possible solution is to obtain material from the "type locality", the site where the original material had been sampled. However, this can be complicated as present results show that multiple cryptic species can occur in sympatry (e.g., three species of *C. curvisetus* at the LTER MareChiara, two species of *C. socialis* complex at the Chilean sample sites.

In the case of sympatric cryptic or pseudo-cryptic entities, several strains belonging to each of them were sampled at different times during this study. I did not encounter any hybrid sequence (species), suggesting no genome exchange between the entities. This leads to conclude that the genetically distinct strains belong to biologically distinct species. For genetically distinct but closely related strains that have been sampled at distant sites, a detailed study is required before they can be considered distinct species, as the genetic differences may simply reflect intraspecific geographic variation. This can be assessed by conducting mating experiments between those geographic strains. If the two produce viable and fecund progeny, then the distinct strains belong to one and the same species, if not, no conclusion can be drawn because of the strains may not have been "in the reproductive mood." Other approaches that can be used to delimit species include physiological studies (as in Degerlund *et al.* 2012) or functional metabolomics (as in Huseby *et al.* 2012).

High-throughput sequencing (HTS) metabarcoding can provide insights in species delimitation. In the HTS analysis, not only the dominant haplotype is shown, but also all the minor ones occurring in the rDNA cistrons. If the reference (Sanger) sequence of a geographically distant specimen groups inside the terminal haplotype clade or cluster of a local entity, then the specimen probably belongs to the local species (see Chapter 4), whereas recovery well outside such a clade or cluster suggests, though does not prove, that the entities belong to biologically separate species.

Morphological evolution: the acquisition sequence of morphological character states can be interpreted by mapping these on the phylogeny inferred from the 18S and partial 28S rDNA gene regions. A couple of examples are illustrated here,

Solitary cells versus colonies: the growth form of members of the family Chaetocerotaceae can be divided into two character states, "solitary" and "colonial - chain forming." The molecular phylogenies revealed that "chain forming" is an ancestral character state of the

family, which is also shared with its close outgroups. This character is not a clear-cut binary one; most species begin their vegetative part of the life cycle as colonies, and then become single-celled later in their life cycle. In some species, this transition from "chain forming" to "solitary" takes place near the beginning of the vegetative life cycle, whereas in others this happens only towards the end of it. Examples of species making this transition in or at the beginning of their vegetative life cycle are *Bacteriastrum parallelum*, *Chaetoceros calcitrans, C. tenuissimus, C. danicus,* and *C. peruvianus.* At times, I observed large cells of the latter two species -probably in the beginning of their vegetative life cycle- as small chains consisting of two or three cells. However, I never observed *C. tenuissimus* as colonial form. In general, the five species mentioned here are found in different clades of the molecular phylogeny and therefore, being "solitary early in the life cycle" is a homoplasy. The transition from "colonial" to "solitary" is not always one-way. For example, during my study, I isolated *C*. cf. *affinis*-V (Na26B1) as single cells, but it later formed colonies composed of cells of a similar size.

Chain morphology: as stated above, most species of the family Chaetocerotaceae are known to form chains. The number of cells in the chains may vary from a few cells to many depending on the species. Chain morphology is represented by the following morphological characters "straight or bent" vs "twisted or torsion" vs "curved or helical." However, from the phylogenetic tree, it cannot be stated which of these is an ancestral state or a derived state, as the related outgroups show both "straight" and "curved" state. Although, from the molecular phylogenies, it's evident that most of the species of the family present "straight or bent" chains, while few species reveal a different character state. Some species that reveal a different character state include the "curved" appearance of chain morphology. Curved chains were developed independently in different species, but the degree was not homogeneous. For example, in the paraphyletic *C. curvisetus* complex, all species reveal a "curved" chain in narrow girdle view, which was a unique derived state as per the molecular

phylogenies (i.e., autapomorphy). The members of *C. socialis* complex shows "curved" chain morphology in broad girdle view, but the mode of curviness is not identical to *C. curvisetus* complex. Similarly, the character state "torsion" in the chain was gained only once (e.g., *C. tortissimus*). My result corroborates with the finding of Rines & Theriot (2003) and Kooistra *et al.* (2010), where the curved chains in *C. socialis* are classified as a different character state. Furthermore, they tend to form a secondary colony, which is not seen in any of the species in Chaetocerotaceae.

Chloroplasts number: the number of chloroplasts in the family can be; "one", "two" or "many." The molecular phylogenies reveal that "many chloroplasts" constitutes an ancestral character state, which is also shared with many outgroups. The phylogenetic tree forms five major clades, with two out of five comprising of species possessing "many" chloroplasts, i.e., *Bacteriastrum* and *C. contortus-compressus*. The rest of the three clades also include representative groups showing "many" chloroplasts. In general, few of the groups among the three clades possess species with "single" chloroplast. These species are found in different parts in the phylogenies, inferring that the "single" chloroplast is a homoplasy. Apart from the "single" chloroplast, another character state with "two" chloroplasts is observed in *C. anastomosans*, *C.* 'brush'*, C. didymus, C. protuberans* and *C. vixvisibilis*, suggesting this character state is a homoplasy according to the molecular phylogenies. In the present case, present result deviates from the previous results of Kooistra *et al.* (2010), where they state that 'two' chloroplast was a unique derived state (i.e., autapomorphy), because of the inclusion of several new species that were not characterised by their work. Hence, with the addition of several other species from the family may result into the better illustration on the evolution of morphological traits.

The presence of introns: during the present investigation, a surprising finding was the encounter of several introns or intragenic sequences in the rDNA genes of Chaetocerotaceae. These introns are present on specific sites, in both partial 28S and 18S rDNA. However,

these introns do not alter the activity of the gene, as they are removed by RNA splicing during the maturation of the final RNA product (Berg *et al.* 2007; Alberts 2008). These introns were known to occur in a few members of Chaetocerotaceae, which are phylogenetically related, i.e., introns were present in the species belonging to two clades. The species harbouring introns in 18S are *B. hyalinum*, *B*. cf. *furcatum,* and *B. jadranum*, which are all found in the *Bacteriastrum*-clade. The other group of related species includes *C. anastomosans*, *C. '*brush*'*, *C*. cf. *pseudocrinitus*, *C. circinalis*, *C. decipiens*, *C. diversus*, *C. rotosporus*, *C. seiracantus*, *C. '*singlecells', *C. '*verylongsetae' and *C. vixvisibilis*. In the partial 28S rDNA, they were observed in a few members of the latter group, namely *C. '*brush*'*, *C*. cf. *pseudocrinitus C. diversus, C.* 'singlecells' and *C*. 'verylongsetae'. Although, from the study on the phylogeny of introns, I do not find any corroboration between their phylogenetic relationships and that of their Chaetocerotacean "host" sequences. Detailed information on the position of the introns, its relationship as compared to the phylogeny of Chaetocerotaceae, its characterisation, and secondary structure is investigated in Chapter 6.

The number of introns is variable in the rDNA gene, but up to eight are present in *C. diversus*. Therefore, increasing product size that in some cases about twice the normal gene size, poses problems in amplification of the gene region. Some of these introns are also present within the primer sites, preventing/hampering/inhibiting either amplification- or sequencing primers, for e.g., the primer target sites of Ch-690F/R, Ch-1147R, and Ch1400R exhibit introns. This may be one of the reasons why this important group of species in the plankton worldwide is under-represented in the 18S rDNA databases.

Another issue regarding introns is their presence in the V4-region of 18S rDNA. In the present study, I found an intron of ca. 130 bp within the V4 region in the strain *C. '*verylongsetae*'* Na13C2 and in all strains of *C. diversus*. In the metabarcoding approach, where amplicons are generated using a marker region such as V4 region, it will be interesting to see if these species are represented or not. The amplicons representing these species will

be generated in this approach, but due to the presence of intron, the amplicons will fail to generated contigs, which may lead to loss of this species in biodiversity assessment. Hence, care needs to be taken to understand the genetic composition of the species, prior to biodiversity assessment using metabarcoding approach.

The V4-region as a marker of choice in high-throughput sequencing metabarcoding of environmental DNA to assess Chaetocerotacean diversity: can the V4 region be used to discriminate the various Chaetocerotacean species? In the last decade, there is a shift from classical LM methods to culture-free, amplicon-based metabarcoding methods to monitor biodiversity. Metabarcodes are generated using High-Throughput Sequencing (HTS). Specific barcode markers in environmental DNA (eDNA) are PCR-amplified and the PCR products sequenced to generate amplicon libraries. The amplicons are sorted into metabarcode haplotypes, and the haplotypes grouped into clusters or terminal clades, i.e., clades without any clear internal resolution. Finally, the obtained clusters or terminal clades are keyed out using reference barcodes i.e., the ones obtained in this study. The results of the present study have demonstrated that the metabarcoding of Chaetocerotacean diversity at the GoN is likely to function because the universal V4 primers anneal with their intended target regions in the 18S rDNA of all but one of the Cheatocerotacean species, namely *C. radicans*, which has never been reported at the GoN. The PCR-product is ca. 400 bp in all of the species. Several Chaetocerotacean species show inserts in their 18S sequences, but only two species exhibit a ca. 110 bp insert in the V4 region. Moreover, all the species identified in the present study have unique V4 regions, i.e., the V4 can be used as reference barcode to distinguish the species. Finally, the fact that the Chaetocerotacean V4 reference sequences from a clade imply that the entire set of them can be used as a query to extract all and only the Chaetocerotacean metabarcodes from a dataset generated from environmental DNA samples collected at the GoN.

Conclusions and outlook

In this chapter, it has been demonstrated that the traditional species delimitation in the family Chaetocerotaceae based exclusively on morphological characteristics, does not fully capture the species diversity. The combination of genetic and morphological information of the same strains aided species delimitation. The large number of genetically distinct strains collected at the different collection sites allowed to resolve phylogenetic relationships within the family in far larger detail that previous studies (Kooistra *et al.* 2010; Li *et al.* 2013, 2015; Bosak *et al.* 2015). The results confirm reportings in Guiry and Guiry (2016) that, at least in the GoN, *Chaetoceros* is highly diverse whereas *Bacteriastrum* is far less diverse. It may seem that the Neapolitan site includes far more species than the sites at Roscoff and Central Chile. However, in my study, the sites at Roscoff and at Concepcion were sampled only once and the sites near Las Cruces a few times in the spring. Instead, the GoN was sampled extensively throughout the seasonal cycle over several years. Therefore, a comparison of the species diversity at these different sites cannot be done based on the results presented here. Nonetheless, comparison of the species found at these sites shows that several species found in the GoN also occur in Central Chile and/or Roscoff. Yet, there are species in Chile that have not been detected in the extensively investigated GoN, suggesting that not everything is everywhere. However, the culture-based techniques on which this conclusion is based may have missed these species in the GoN. This is why I proceed in the next chapter to use a culture-free approach for assessing the Chaetocerotacean diversity in plankton samples, namely HTS metabarcoding. The 18S rDNA sequences of the various Chaetocerotacean species generated in this Chapter are used as reference barcodes to assess species diversity and seasonal changes in this family at the LTER MareChiara station in the Gulf of Naples (Chapter 4) by means of an HTS metabarcoding approach.

Species diversity of Chaetocerotaceae in the Gulf of Naples as revealed by High Throughput Sequencing - meta-barcoding

Abstract

The present study focuses on assessing the species diversity in the family Chaetocerotaceae in the Gulf of Naples, Italy, using a high-throughput sequencing (HTS) metabarcoding approach. During 2011-2013 a total of 48 surface seawater samples collected at the LTER MareChiara were processed for High-throughput Sequencing (HTS) metabarcoding of the V4 hypervariable region in the 18S rDNA. The 13.6M obtained contigs grouped into 615,142 metabarcode haplotypes. A total of 18,625 of these metabarcode haplotypes were assigned to Chaetocerotaceae, based on EPA and BLAST analysis using V4 sequences of taxonomically validated Chaetocerotacean strains as reference barcodes. Of these, only 650 were retained following the elimination of all metabarcode haplotypes with less than three contigs. The retained metabarcode haplotypes were aligned with the reference barcodes to build an ML tree. The inferred tree resolved 66 terminal clades assigned to *Chaetoceros* and ten to *Bacteriastrum*. Of these 76 terminal clades, 39 included at least a Neapolitan reference barcode, i.e., the clade was assigned to that species. Seven terminal clades contained at least a reference barcode from a strain sampled elsewhere, from a species not known to occur in the Gulf of Naples. The remaining 30 terminal clades did not include a reference barcode at all, meaning that these species are unknown. Within *Bacteristrum*, five terminal clades were represented by a reference barcode. A clade with a reference barcode of a strain belonging to *B. hyalinum* sampled at the LTER Marechiara was not associated to any HTS-metabarcode, and four metabarcode clades had no reference barcodes. Thus, the HTS-metabarcoding approach combined with the taxonomic exploration of the Chaetocerotacean diversity in the Gulf described in the previous chapter has uncovered a considerable number of species not previously known in our study area.

Introduction

High species diversity enhances ecosystem complexity, resilience against disturbances and adaptability to environmental change. Therefore, accurate knowledge of the biodiversity in an ecosystem is a prerequisite towards better predictions on how these ecosystems will respond to change. With regard to marine phytoplankton, species diversity has been explored in earnest over at least one and a half century. Various tools that have become available over time (e.g., phase contrast and confocal light microscopy, electron microscopy, sequencing or fingerprinting of DNA markers, flow cytometry, and flow-cam) have diversified the ways and increased the details in which these microalgae can be studied, and even changed the views of what defines species in the phytoplankton.

Recently, high-throughput sequencing (HTS) techniques of universal, taxonomically informative marker regions –so called metabarcodes- has been applied successfully to assess the overall biodiversity of various protistan and bacterial communities. Several studies have now assessed biodiversity in samples taken from soil (Bougoure & Cairney 2005; Cong *et al.* 2015; Orgiazzi *et al.* 2015), from aquatic (Graham *et al.* 2004; Zinger *et al.* 2012; Eiler *et al.* 2013; Wang *et al.* 2014; Kammerlander *et al.* 2015; Visco *et al.* 2015), marine benthic (Takishita *et al.* 2007; Fonseca *et al.* 2014; Massana *et al.* 2015) and planktonic (Savin *et al.* 2004; Amaral-Zettler *et al.* 2009; Andersson *et al.* 2010; Behnke *et al.* 2010; Edgcomb *et al.* 2011; Lindeque *et al.* 2013; Decelle *et al.* 2014; Massana *et al.* 2004, 2011, 2014, 2015; Zimmermann *et al.* 2014; Malviya *et al.* 2016; Piredda *et al.* 2017) habitats, as well as from, for instance, the gut content of animals (Boyer *et al.* 2013; Kartzinel & Pringle 2015; Lyons *et al.* 2015) and humans (Reyes *et al.* 2012; Ji & Nielsen 2015). All of these studies showed higher biodiversity than expected based on culture-based methods.

Monitoring of all this biodiversity at Long Term Ecological Research (LTER) sites has provided insight into its changes over different timescales. The collected information enables

assessment of correlations among environmental factors and biodiversity data in ever-increasing detail, and hence formulation of testable hypotheses on causal interrelationships between species (co)-occurrence and various biotic and abiotic factors. Results of these efforts can then provide more realistic input in modeling exercises to predict global change.

In spite of all this progress, knowledge of species diversity in the plankton is far from complete. Recent taxonomic studies have demonstrated that even such morphologically conspicuous and well-studied planktonic diatom genera such as *Chaetoceros* (Kooistra *et al.* 2010; Chamnansinp *et al.* 2013, 2015), *Pseudo-nitzschia* (Amato *et al.* 2007; Amato & Montresor 2008; Churro *et al.* 2009; Quijano-Scheggia *et al.* 2009; Lim *et al.* 2012; Lundholm *et al.* 2012) and *Skeletonema* (Sarno *et al.* 2005, 2007; Kooistra *et al.* 2008) include many species new to science. This result contrasts with the findings of Nanjappa *et al.* (2013, 2014) revealing low diversity in the widely distributed planktonic diatom genus *Leptocylindrus*.

In the previous chapter, exploration of species diversity in the family Chaetocerotaceae at the LTER MareChiara (LTER-MC) in the Gulf of Naples (GoN) and elsewhere, resulted in an 18S rDNA reference dataset of 77 species (40 from GoN only); nine of which were new to science and three were known from elsewhere, but never recorded before in the GoN. The overarching objective of the High-throughput Sequencing (HTS) metabarcoding study presented here is to evaluate the effectiveness of aforementioned efforts 1) to assess the Chaetocerotacean diversity at the LTER-MC and, 2) to validate if metabarcoding approach reveals the entire diversity in the family. HTS-metabarcoding of the entire eukaryotic community allows detection of all terminal clades in a sample, assuming that these clades represent the species in that sample. Comparison with reference barcodes of Chaetocerotacean species gathered in Chapter 3 and from existing reference databases (PR2, Guillou *et al.* 2013 and SILVA, Quast *et al.* 2013; Yilmaz *et al.* 2014) reveals which clades

represent which species and which clades belong to species still unknown to us. To uncover seasonal cycles and interannual variation in this diversity, I explored changes in the diversity of the Chaetocerotacean metabarcodes along 48 plankton samples gathered over three consecutive years (2011-2013). To assess if cell counts in light microscopy corroborate the HTS-metabarcode data, I compared the two types of data gathered from the samples.

The 18S rRNA gene includes nine variable regions (V1 to V9; Nelles *et al.* 1984; Neefs *et al.* 1993; Van De Peer & Wachter 1997; Wuyts *et al.* 2001; Ki 2012). Among these, the V2, V4, and V9 are considered be most suitable for biodiversity assessments, the V4 region being the most frequently used (Nickrent & Sargent 1991; Hadziavdic *et al.* 2014). The V4-region in the 18S rRNA-gene was selected as marker of choice, since it constitutes the longest and arguably most variable region among them. Since its length usually does not exceed the ca. 400 bp, it can easily be generated using the Illumina MiSeq platform.

Material and Methods

The study site, the Long Term Ecological Research station Mare-Chiara (LTER-MC) in the Gulf of Naples, Italy, has been investigated extensively for biological and physical parameters since 1984. The station is located in the transition zone between coastal, nutrient-rich water and the oligotrophic water of the central Tyrrhenian Sea. A sampling of physical and biological parameters at LTER-MC is conducted weekly. Since 2010, sampling at the study site has been complemented with a collection of filtered material for environmental DNA extraction and metabarcode analysis.

Sample collection for HTS analysis and genomic DNA extraction: every week, 3 L of surface seawater was collected and three biological replicates were filtered on a cellulose ester filter (47 mm diameter, 1.2 μm pore size; EMD Millipore, Billerica MA, USA). Filters were frozen immediately in liquid nitrogen and stored at -80 °C until further processing. Among these, 48 filters selected to track the diversity, abundance, and seasonal trend of

various species from the year 2011-2013. For each temporal sample, genomic DNA was extracted from filters using QIAGEN DNeasy 96 Plant kit (QIAGEN GmbH, Hilden, Germany), following manufacturer's instructions by Dr. Roberta Piredda and Dr. Maria Paola Tomasino, in Stazione Zoologica Napoli (SZN). DNA concentration and quality were measured with a NanoDrop 1000 (Thermo Fisher Scientific Inc., Wilmington DE, USA).

Morphological enumeration in phytoplankton samples: A detailed description is presented in Chapter 2.

HTS analysis: the Universal V4 primers were used to amplify the V4 region of the 18S rDNA gene. The HTS procedure was carried out at the Molecular Biodiversity Lab (MoBiLab) of the ESFRI LifeWatch-Italy (Bari, Italy). The Illumina Nextera's protocol (Illumina, San Diego, CA, USA) was modified to prepare the 18S rDNA-V4 amplicon library, to be sequenced on the Illumina MiSeq platform (Kozich *et al.* 2013; Manzari *et al.* 2015).

PCR amplification and HTS sequencing: the protocol is based on two-step amplification: the first amplification is carried out with V4 universal primers with a 5'-end-specific overhang (adapter, indicated in small letters). The V4- primers (capital boldface) along with the adapters read as follows: V4_18SNext forward 5'-tcg tcg gca gcg tca gat gtg tat aag aga cag **CCA GCA SCY GCG GTA ATT CC**-3' and V4_18SNext reverse as 5'-gtc tcg tgg gct cgg aga tgt gta taa gag aca g**AC TTT CGT TCT TGA TYR ATG A-**3'.

Two separate amplifications were performed for the two half filters for each sample date, and the PCR products were pooled in one sample per date. PCR amplifications for the V4 region (PCR I) were performed each in 25 µl reaction mixture containing 2.5 ng or 5.0 ng of extracted DNA, 1X Buffer HF, 0.2 mM dNTPs, 0.5 µM of each primer, and 1U of Phusion High-Fidelity DNA polymerase (New England Biolabs Inc, Ipswich, MA, USA).The PCR cycles were standardized as follows: initial denaturation at 98 °C for 30 sec, followed by 10

cycles of denaturation at 98 °C for 10 sec, annealing at 44 °C for 30 sec, extension at 72 °C for 15 sec, and subsequently 15 cycles of denaturation at 98 °C for 10 sec, annealing at 62 °C for 30 sec, extension at 72 °C for 15 sec, with a final extension step of 7 min at 72 °C. All PCR's were performed in the presence of a negative control (RNase/DNase-free water). The PCR products (ca. 470 bp for V4) were visualized on 1.2% agarose gel and purified using the AMPure XP Beads (Agencourt Bioscience Corporation, Beverly, MA, USA), at a concentration of 1.2X vol/vol, according to manufacturer's instructions.

The purified PCR products served as a template for a second amplification (PCR II) using a mixture of the Nextera index primers and Illumina P5 and P7 primers as required by the Nextera dual index procedure. This procedure included incorporation of unique indices in both ends of the library molecules, allowing sample identification for bioinformatics analysis (Kozich *et al.* 2013). The PCR II included a 50-µl reaction mixture included: template DNA (40 ng from PCR I), 1X Buffer HF, dNTPs (0.1 mM), primers (Nextera index primer 1 and 2; Illumina primers P5 and P7) and 1U Phusion DNA Polymerase. The cycling parameters were those suggested by the standard Illumina Nextera protocol. The final amplicon size is ca. 550 bp (including ca. 400 bp of V4 region and 150 bp of Illumina Nextera adapters) was purified using AMPure XP Beads, at a concentration of 0.6X vol/vol. The quality of the product was analysed on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and quantified by fluorimetry using the Quant-iTTM PicoGreen-dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) on NanoDrop 3300 (Thermo Fisher Scientific). Equimolar quantities of V4 amplicons were pooled and subjected to 2x 250 bp sequencing on a MiSeq platform to obtain a total of about 375,000 pair-end reads per sample.

Processing of the raw HTS data into contigs: the sequences generated from the Illumina platform was assessed for sequence quality using the FastQC tool on the Galaxy platform (http://usegalaxy.org/) before assembling. Illumina paired-end reads (ca. 250bp) were

processed using Mothur v1.33.0 (Schloss *et al.* 2009), using the standard operating procedure (http://www.mothur.org/wiki/MiSeq_SOP; Kozich *et al.* 2013). Our initial HTS raw data included 48 surface water samples and two sediment samples. Sediment counts were excluded, as these were not relevant to my objectives. Pairs of forward and reverse sequences in the raw HTS data were aligned and merged to generate contigs. Overlap was on average 81 bp (Stdev 11.3). Differences in base calls in the overlapping region were solved using the ΔQ parameter as described in Kozich *et al.* (2013). A flowchart stating the different steps in pre-processing of the metabarcoding data to the final dataset is presented in Fig. 4.1.

Cleaning and de-replication of contigs: the contigs were screened to fit within a length between 390-440 bp and those not matching this criterion were removed. Then, the primer sites were removed (pdiffs=3). Contigs were inspected for the presence of homopolymers and those longer than 8 bp were removed. Contigs with any ambiguity in their overlapping part were eliminated as well. The cleaned contigs were sorted into groups of identical sequences.

Removal of non-target sequences and contigs belonging to bacteria, fungi, and metazoa: representative contigs were aligned to the reference alignment SILVA SEED v119 (Quast *et al.* 2013; Yilmaz *et al.* 2014). Any representative contig that did not align with any of the reference V4-region was eliminated. Aligned representative contigs were classified using a naive Bayesian classifier (Wang *et al.* 2007), using the PR2 database (Guillou *et al.* 2013) as a training set with an 80% bootstrap confidence threshold. Representatives assigned to Eubacteria, Fungi or Metazoa were removed.

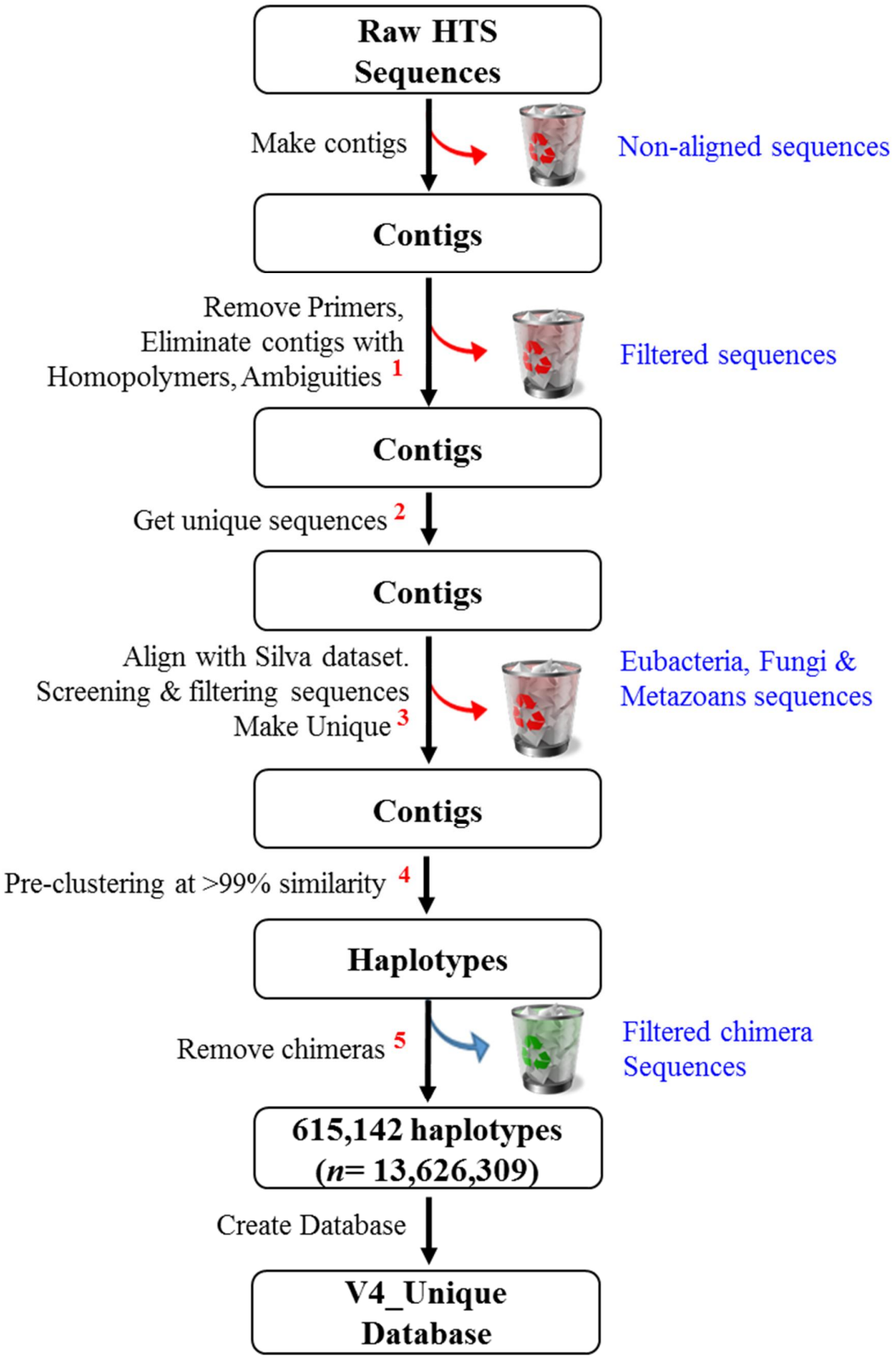
Pre-clustering of the representatives: a pre-clustering algorithm (Huse *et al.* 2010) was used to further de-noise representatives; allowing one nucleotide difference for every 100 bp (99%) of the sequence. Pre-clustered representatives and their associated identical contigs are from here onwards referred to as "*metabarcode haplotypes*."

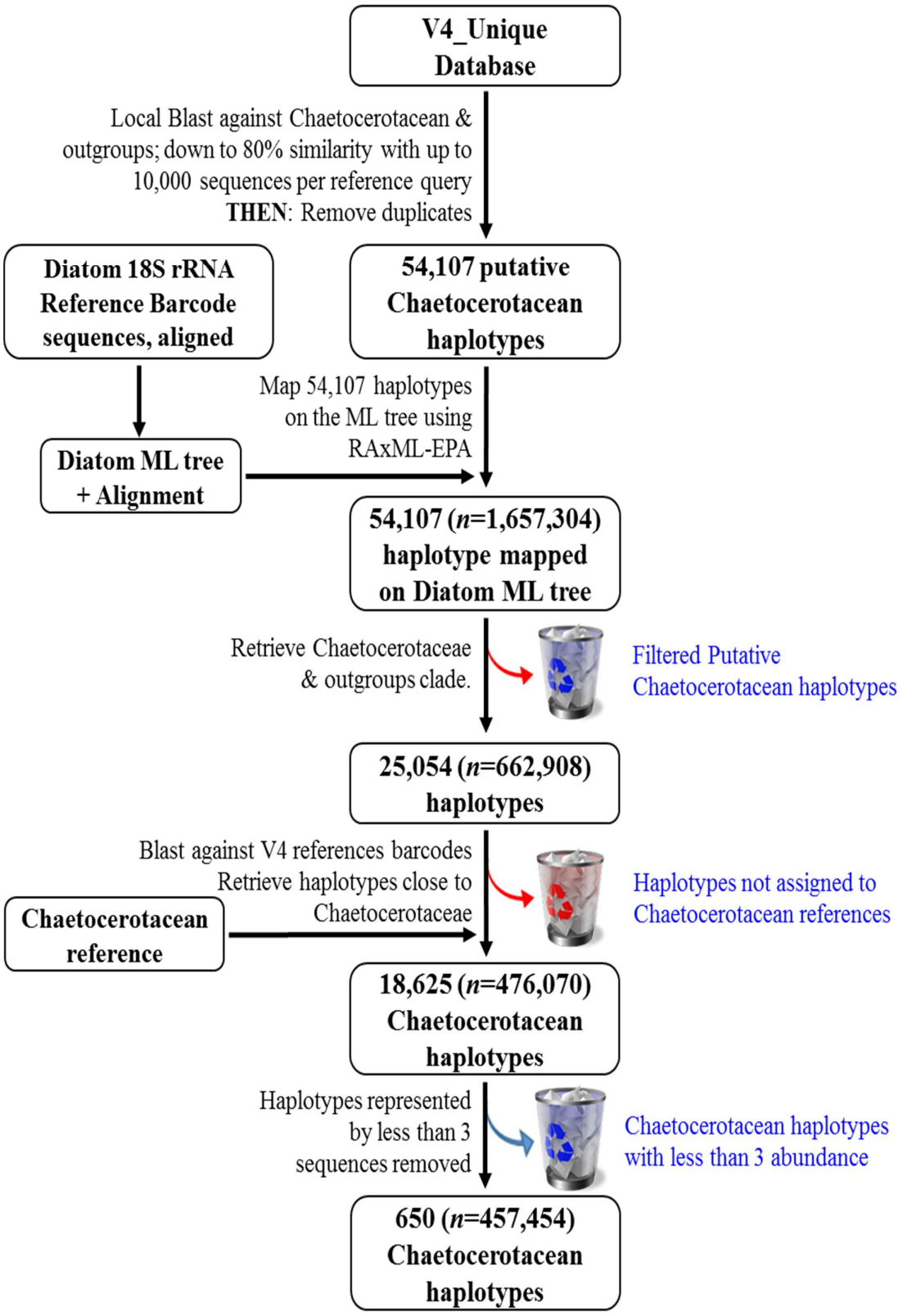
Removal of chimeras: the metabarcode haplotypes were screened for chimeras using UCHIME (Edgar *et al.* 2011) de novo mode in Mothur; chimeras were removed.

The final dataset contained a total of 13,626,039 contigs, grouped into 615,142 metabarcode haplotypes. A count table was generated in Mothur, listing the abundance of contigs for each metabarcode haplotype (in the rows) for each sample (in the columns). A local database of these 615,142 fasta haplotypes was generated using the formatdb command in BLAST package from the NCBI.

Gathering of Chaetocerotacean metabarcode haplotypes:

Selection of putative Chaetocerotacean metabarcode haplotypes: a V4 dataset of 149 reference barcodes was generated. The V4 reference barcodes included 121 Chaetocerotacean barcodes, i.e., 77 taxonomically validated references (from known Chaetocerotacean species and from cultures of Chaetocerotacean species new to science) as well as 44 environmental sequences of unknown Chaetocerotacean species and 28 sequences of near outgroups from GenBank. The 121 Chaetoceroatacean reference barcodes were used as queries to perform a local BLAST against the local database of 615,142 metabarcode haplotypes. Metabarcode haplotypes were retrieved down to either a threshold of 80% similarity match OR a query output maximum of 10,000 of them (Massana *et al.* 2014), whatever threshold was reached first, using the string command -perc_identity 80 –max_target_seqs_10,000. Since in this way, numerous metabarcode haplotypes were retrieved by many query sequences, the multiple hits were filtered away from the dataset of putative Chaetocerotacean metabarcode haplotypes. The procedure selected 54,107 unique, putative Chaetocerotacean haplotypes (Fig. 4.1).


Raw HTS Sequences
Make contigs
Non-aligned sequences
Contigs
Remove Primers, Eliminate contigs with Homopolymers, Ambiguities 1
Filtered sequences
Contigs
Get unique sequences 2
Contigs
Align with Silva dataset. Screening & filtering sequences Make Unique 3
Eubacteria, Fungi & Metazoans sequences
Contigs
Pre-clustering at >99% similarity 4
Haplotypes
Remove chimeras 5
Filtered chimera Sequences
615,142 haplotypes (n= 13,626,309)
Create Database
V4_Unique Database


V4_Unique Database
Local Blast against Chaetocerotacean & outgroups; down to 80% similarity with up to 10,000 sequences per reference query
THEN: Remove duplicates
Diatom 18S rRNA Reference Barcode sequences, aligned
54,107 putative Chaetocerotacean haplotypes
Map 54,107 haplotypes on the ML tree using RAxML-EPA
Diatom ML tree + Alignment
54,107 (n=1,657,304) haplotype mapped on Diatom ML tree
Retrieve Chaetocerotaceae & outgroups clade.
Filtered Putative Chaetocerotacean haplotypes
25,054 (n=662,908) haplotypes
Blast against V4 references barcodes
Retrieve haplotypes close to Chaetocerotaceae
Chaetocerotacean reference
Haplotypes not assigned to Chaetocerotacean references
18,625 (n=476,070) Chaetocerotacean haplotypes
Haplotypes represented by less than 3 sequences removed
Chaetocerotacean haplotypes with less than 3 abundance
650 (n=457,454) Chaetocerotacean haplotypes

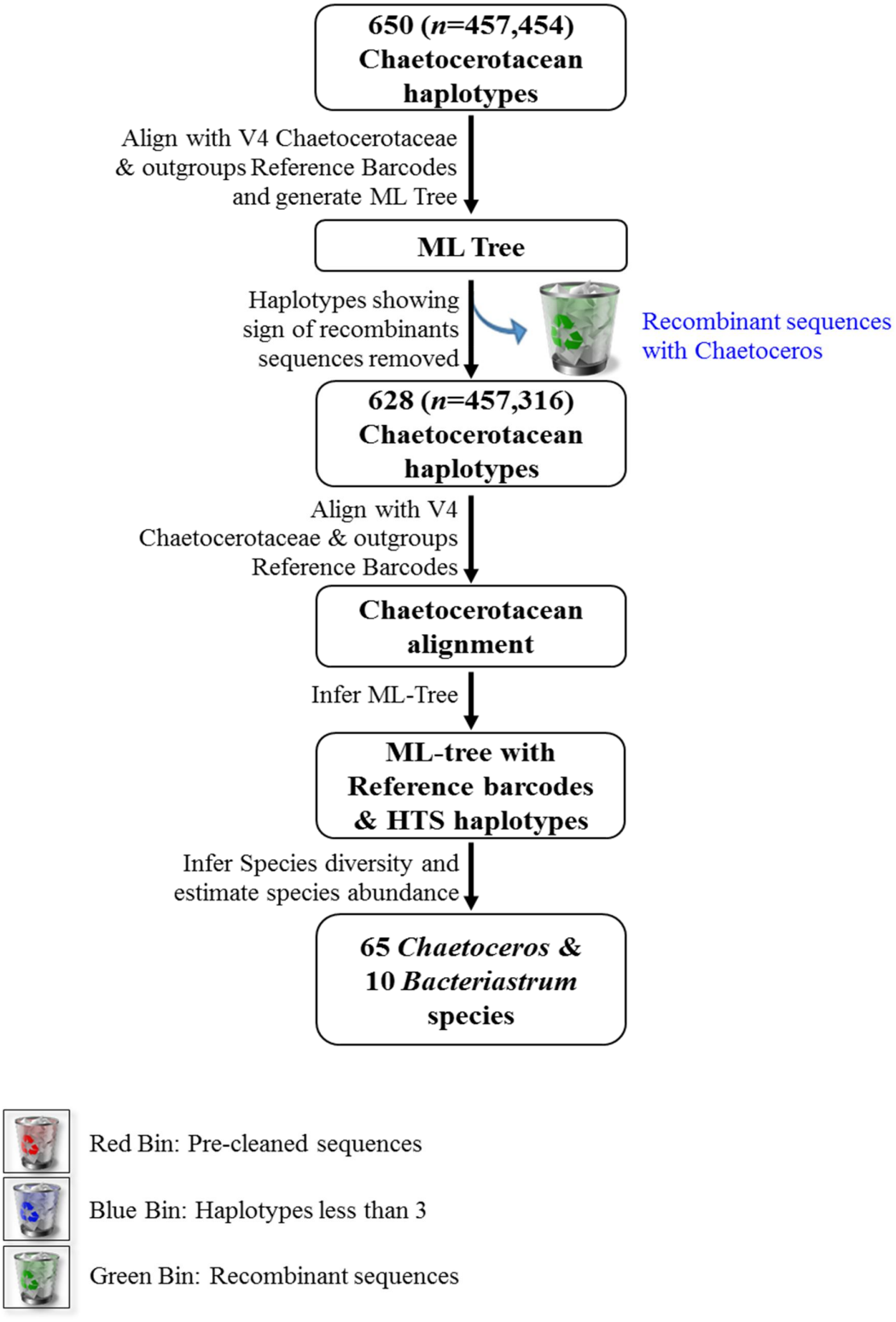


Fig. 4.1: A flowchart showing the pre-processing steps, collection of the Chaetocerotacean metabarcode haplotypes from the HTS data and inferring the diversity by constructing a maximum likelihood tree using FastTree. Preprocessing steps (indicated as 1-5 in fig) are discussed in the materials and methods in detail on p. 121-122 and p. 125-126.

Validation of the putative Chaetocerotacean metabarcode haplotypes using an 18S rDNA phylogeny of the diatoms: a diatom 18S rDNA reference dataset was generated with full-length sequences of Chaetocerotacean species as well as outgroups from across the centric and pennate diatoms. This reference dataset comprised 514 unique sequences. The sequences were aligned using MAFFT v7.245 (Katho *et al.* 2002). Minor adjustments were made to the alignment by eyeball in SeaView v4.5.4 (Gouy *et al.* 2010). A maximum likelihood (ML) tree was generated with this alignment using RAxML Blackbox (http://embnet.vital-it.ch/raxml-bb/; Stamatakis *et al.* 2008) under default conditions, using a GAMMA+P-Invar model of rate heterogeneity in which all parameters were estimated. Support for internal internodes was estimated by means of 100 rapid bootstrap replicates. To select all –and only– the Chaetocerotacean metabarcode haplotypes from the *putative* Chaetocerotacean metabarcode haplotypes, the latter were mapped over the above ML-phylogeny using the RAxML Evolutionary Placement Algorithm (EPA) on a RAxML stand-alone offline platform (Berger *et al.* 2011). A total of 25,054 metabarcode haplotypes (representing 662,908 contigs) mapped onto the Chaetocerotacean clade (Fig. 4.1). These metabarcode haplotypes were used as queries in a local BLAST against the sequences in the diatom 18S rDNA reference dataset. All metabarcode haplotypes assigned to Chaetocerotaceae were retained in a new dataset, and a count table was generated for these metabarcode haplotypes. The count table represented the occurrence -and if yes, abundance- of each metabarcode haplotype in each sample. A heatmap showing the temporal distribution and occurance is illustrated in Fig. 4.4. Subsequently, metabarcode haplotypes representing less than three contigs were eliminated from the dataset. The final Chaetocerotacean dataset included 650 metabarcode haplotypes **(**representing 457,454 contigs**)**.

Tree construction with metabarcode haplotypes and V4-reference barcodes: the obtained Chaetocerotacean metabarcode haplotypes were aligned with the 148 V4-reference barcodes of Chaetocerotaceae and near outgroups using MAFFT v7.245 (Katho *et al.* 2002).

Minor adjustments were made to this obtained alignment by eyeball in SeaView v4.5.4 (Gouy *et al.* 2010). An ML-tree was inferred from the obtained alignment using FastTree v2.1.8 No SSE3 (Price *et al.* 2009, 2010) and the same parameters as in Chapter 3, i.e., a GTR model with rate heterogeneity. Tree length was reduced with subtree-prune-regraft (SPR) moves -nearest-neighbor interchanges (NNIs), and the significance of the difference in likelihood values between the shorter tree and its "parent" was tested using Shimodaira-Hasegawa (SH) tests.

Results

The dataset and phylogeny of V4-reference barcodes: the 121 Chaetocerotacean V4 reference barcodes averaged ca. 380 bp in length and ranged between 370 bp (*Chaetoceros* clone EU371179) and 386 bp (*C. curvisetus*1 DH8), excluding the primer sites and inserts. The V4 regions of *C.* 'verylongsetae' strain Na13C2 and *C. diversus* included an insert of 130 and 120 bp, respectively, from position 323 and 325 onwards, rendering their regions 510 bp and 500 bp in length, respectively. The V4 barcode of *Bacteriastrum hyalinum* strain CCMP141 exhibited an insert at position 323 while the Neapolitan strains belonging to this species lacked such an insert.

The V4 reference alignment comprises of 121 Chaetocerotacean reference barcodes with the 28 barcodes of close outgroup taxa and spanned 395 positions, excluding primer sites and inserts. For the inference of the V4-tree, a GTR model was chosen with a base substitution rate matrix of A↔C (0.9447), A↔G (2.6737), A↔T (0.8549), C↔G (1.2170), C↔T (4.4340), relative to G↔T (1.0000); a gamma-distributed rate variation among sites with shape parameter $\alpha = 0.3370$ and base frequencies A 0.2653, C 0.1588, G 0.2608, and T 0.3150. The likelihood score (-ln L) of the resulting ML tree was 8,065.333. In this tree, the outgroup sequences formed a reciprocal sister clade to a clade including all

Chaetocerotacean sequences with 95% bootstrap support. The Chaetocerotacean reference barcodes grouped in six distinct clades.

The phylogeny of 18S rDNA sequences: for the inference of the ML phylogeny of 514 aligned diatom 18S rDNA sequences a GTR model was chosen with a base substitution rate matrix of A↔C (0.8944), A↔G (2.3521), A↔T (1.1512), C↔G (0.8195), C↔T (4.0066), relative to G↔T (1.0000); a gamma-distributed rate variation among sites with shape parameter $\alpha = 0.6017$ and base frequencies A 0.2660, C 0.1927, G 0.2599, and T 0.2815. The likelihood score (-ln L) for the ML-tree was 114,761.1963. In this ML-tree, the reference barcodes of *Chaetoceros* grouped in a very weakly resolved clade (51%), whereas those of *Bacteriastrum* (100%) formed a well-supported clade (100%). The two genera were not nearest sisters but grouped in a clade (100%) including other polar centrics. *Dactyliosolen* resolved as the nearest sister of *Chaetoceros*. The next nearest sister clade (100%) included *Acanthoceros* and *Urosolenia*, and then followed by a clade with *Hemiaulus* (100%). Relationships among the four clades were badly supported or not supported at all. The nearest sister clade to this badly supported clade comprised a well-supported clade (91%) with *Guinardia*, *Cerataulina,* and *Eucampia*. Yet again, the support for the entire clade, including all these genera was low (62%). The sister clades to this latter clade comprised two well-supported clades (100% each), one with *Terpsinoe* and *Hydrosera*, and the other one with *Bacteriastrum*.

V4 metabarcode database: the sorting, screening, clustering, and filtering of contigs resulted in a dataset of 615,142 eukaryotic metabarcode haplotypes, representing a total of 13,626,039 contigs. The local BLAST of these haplotypes with the Chaetocerotacean V4-reference barcodes as query and subsequent de-replicating returned 54,107 metabarcode haplotypes *putatively* belonging to Chaetocerotacean species. Mapping of these 54,107 haplotypes over the 18S rDNA ML tree resulted in 25,054 metabarcode haplotypes

(*n*=662,908 contigs) grouping in a clade containing all the reference barcodes of *Chaetoceros, Bacteriastrum,* and the other abovementioned near outgroup genera. All these 25,054 metabarcode haplotypes were taken from this tree. A local Blast was performed using these haplotypes as a query against the Chaetocerotacean V4-reference barcodes along with those belonging to the abovementioned other polar centric diatoms, resulting in 18,625 metabarcode haplotypes (*n*=476,070 contigs) grouping *specifically* with *Chaetoceros* or *Bacteriastrum*. Of these, 650 (*n*=457,454 contigs) remained following removal of haplotypes containing less than three contigs.

These 650 metabarcode haplotypes were aligned with the 149 V4-reference barcodes of Chaetocerotaceae and close outgroups. The ML phylogeny inferred using FastTree uncovered a series of terminal clades, generally exhibiting only weakly resolved internal structure, if any at all. On a few occasions, a secondary terminal clade was recovered on a long branch emerging from within such a terminal clade. Such secondary terminal clades resulted either from alignment errors or from short chimeras at the 5'-end or the 3'-end on some of the metabarcode haplotypes. Following correction of these errors and exclusion of the chimera from the metabarcode haplotypes, the resulting final alignment contained a total of 628 metabarcode haplotypes, representing 457,316 contigs, along with the 148 V4-reference barcodes.

The phylogeny of metabarcode haplotypes and reference barcodes: an ML tree was inferred from the final alignment in FastTree using a GTR model with a base substitution rate matrix of A↔C (0.9484), A↔G (1.3159), A↔T (0.4859), C↔G (0.4996), C↔T (2.1091), relative to G↔T (1.0000); a gamma-distributed rate variation among sites with shape parameter α=0.6470 and base frequencies A 0.2707; C 0.1552; G 0.2556 and T 0.3185. The likelihood score (-ln L) for the ML tree was 16986.744.

The tree (Fig. 4.2) resolved the 28 outgroup sequences in a reciprocal sister clade (71% bootstrap support) to a clade with the Chaetocerotacean reference barcodes and metabarcode haplotypes. The Chaetocerotacean clade included eight principal clades with high bootstrap support onto a series of basal clades with bootstrap support ranging between 43% (insufficient) and 78%.

Chaetoceros teres and *C. lauderi* Clades

The first of these principal clades included in its turn two clades, one (99%) with reference barcodes of *C. teres* and *C. lauderi* and associated metabarcode haplotypes, and the other one (98%) with several different reference barcodes of *C. brevis/C. pseudobrevis*, each with their associated metabarcode haplotypes, all of these belonging within the subgenus *Hyalochaete*. The species *C. teres* was not known from the Gulf of Naples, nevertheless, one metabarcode haplotype (*n*=11 contigs), detected in March and May of 2011, was identical to a Chilean reference barcode of this species. Within *C. lauderi,* the Neapolitan reference barcodes were close or identical to one co-dominant metabarcode haplotype (*n*=456) encountered from October to March. The other co-dominant haplotype in this terminal clade (*n*=4,052) was present in March-April and from June to October. During the summer of 2016, I sampled and barcoded additional strains of this species but all of these grouped with the Neapolitan reference barcodes. Within the clade of *C. brevis and C. pseudobrevis,* four clades included metabarcode haplotypes. The first of these was recovered as sister to the reference barcode of Chilean strain Ch9B3 and was found only in Feb. 2012. The second set of metabarcode haplotypes formed a terminal clade including a Neapolitan reference barcode of *C. brevis*. These metabarcodes were recorded on and off throughout the year but were most frequent in winter. The third clade contained a Neapolitan reference barcode (of strain Na7C2), which was identical to a metabarcode haplotype (*n*=176) present from the late autumn to the winters of 2011 and 2013. The fourth clade included a Caribbean reference

barcode (GU823798) next to a dominant metabarcode haplotype (*n*=1,127) occurring during summer and autumn (Sept.-Oct.).

Subgenus *Chaetoceros* clade along with *C. protuberans* and *C. didymus*

The second principal clade (99%), included in its turn a clade (70%) bearing a series of terminal clades, including the reference barcodes belonging to taxa the subgenus *Chaetoceros* as sister to a clade (68%) including in its turn two clades (each 99%), one with the reference barcodes of *C. didymus* associated metabarcode haplotypes, and the other one with those of *C. protuberans,* both of these belonging within the subgenus *Hyalochaete*. Within the subgenus *Chaetoceros*, the reference barcodes of several species not known from the Gulf of Naples were, as expected, not accompanied by any metabarcodes. The reference barcodes of *C. peruvianus* and *C. rostratus* resolved in a clade and a grade, respectively, surrounded each by several metabarcode haplotypes, and so was the reference barcode of *C. danicus*. The metabarcodes of these species were detected year-round but were most prominent in the cool seasons. Within *C. didymus,* the metabarcode haplotypes were seen on and off year round but dominated in autumn. Notably, the Chilean reference barcode of *C. didymus* differed markedly from the Neapolitan one and did not group closely with any metabarcode haplotypes. The metabarcode haplotypes of *C. protuberans* were observed on and off over the year but were most abundant in early spring. Notably, the Neapolitan and Chilean reference barcodes of this species were identical, but differed slightly from that of a strain from Bristol (UK), though the latter barcode was recovered well within the terminal clade of this species.

***Chaetoceros contortus* Clade**

The third principal clade (100%) contained several reference barcodes of *C. contortus* as well as several clades with associated metabarcode haplotypes. One of the terminal clades with metabarcode haplotypes (n=206) observed only in winter, did not group closely with

any of the reference barcodes. Instead, the second dominant haplotype (*n*=1,026), grouped closely with a Neapolitan reference barcode (of strain Na30A1) and was detected throughout the year except in autumn and was most abundant in March.

Bacteriastrum Clade

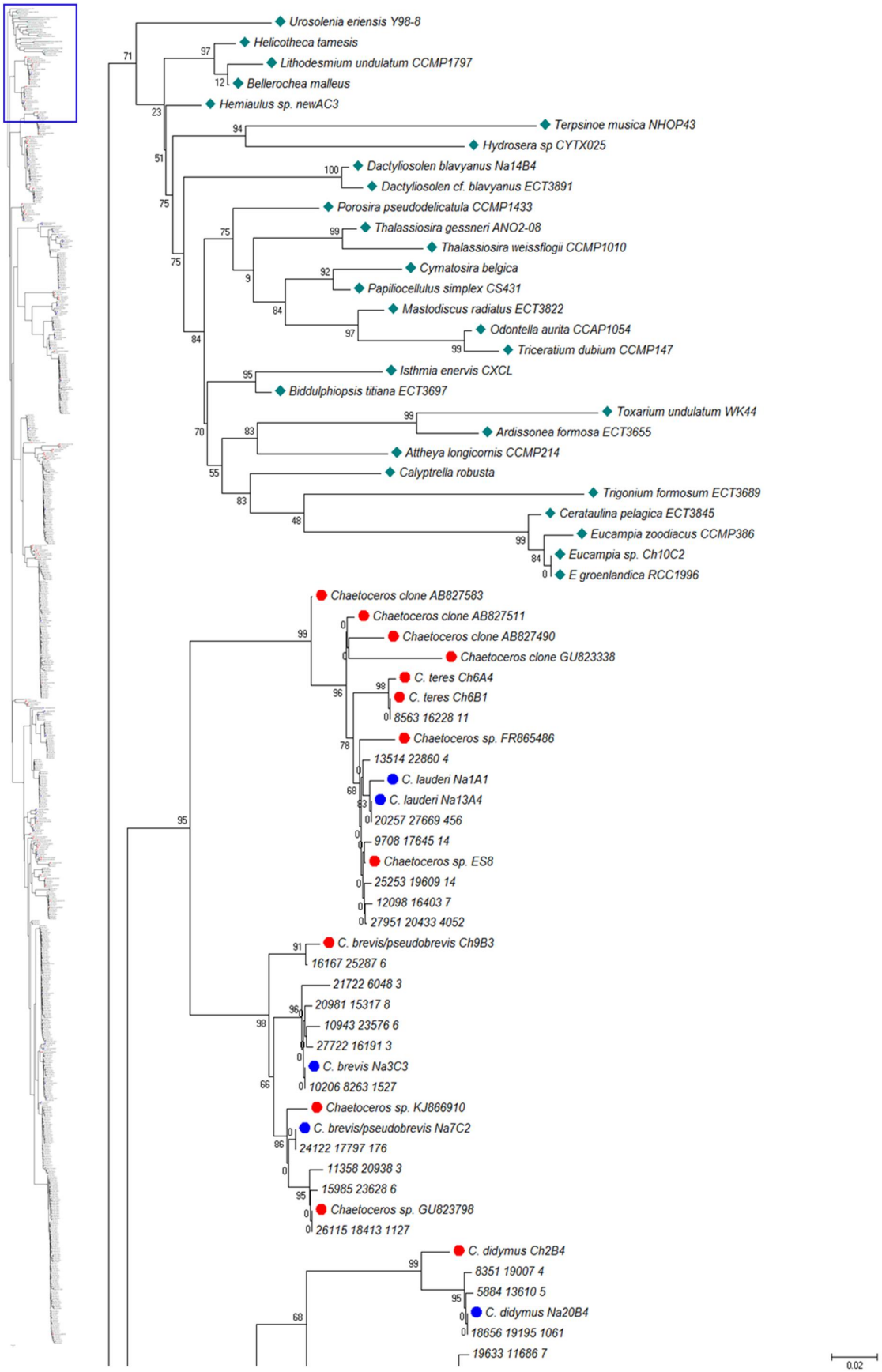
The fourth principal clade (99%) included six terminal clades with the reference barcodes of *Bacteriastrum* and associated metabarcode haplotypes as well as four branches, each with a single haplotype. The clade of *B. hyalinum* contained only two reference barcodes from Neapolitan strains sampled in different years. A BLAST search of one of these references as a query against the raw data confirmed that *B. hyalinum* metabarcodes were indeed absent. The lineage with the metabarcode haplotype 21613_20455 represented 17 contigs, sampled in Feb. 2012 and Nov. 2013. The metabarcode haplotypes of *B. mediterraneum* were present in the autumn. The fourth clade consisted of a single unknown haplotype 17269_14290 (*n*=4) detected in January of 2011 and 2013. Within the clade of *B. elegans* the metabarcode haplotype (*n*=15), which was observed only in March 2012, differed from the reference barcode. The metabarcode haplotypes of *B.* cf. *furcatum* were detected during the winter and those of *B. jadranum* from the autumn into the spring of every year sampled. The two unknown haplotypes were observed in Feb. 2012. *Bacteriastrum parallelum*, with its abundant dominant haplotype (*n*=7,961) and 20 peripheral haplotypes, constituted the most abundant species in this principal clade. It was detected year-round in the metabarcode dataset.

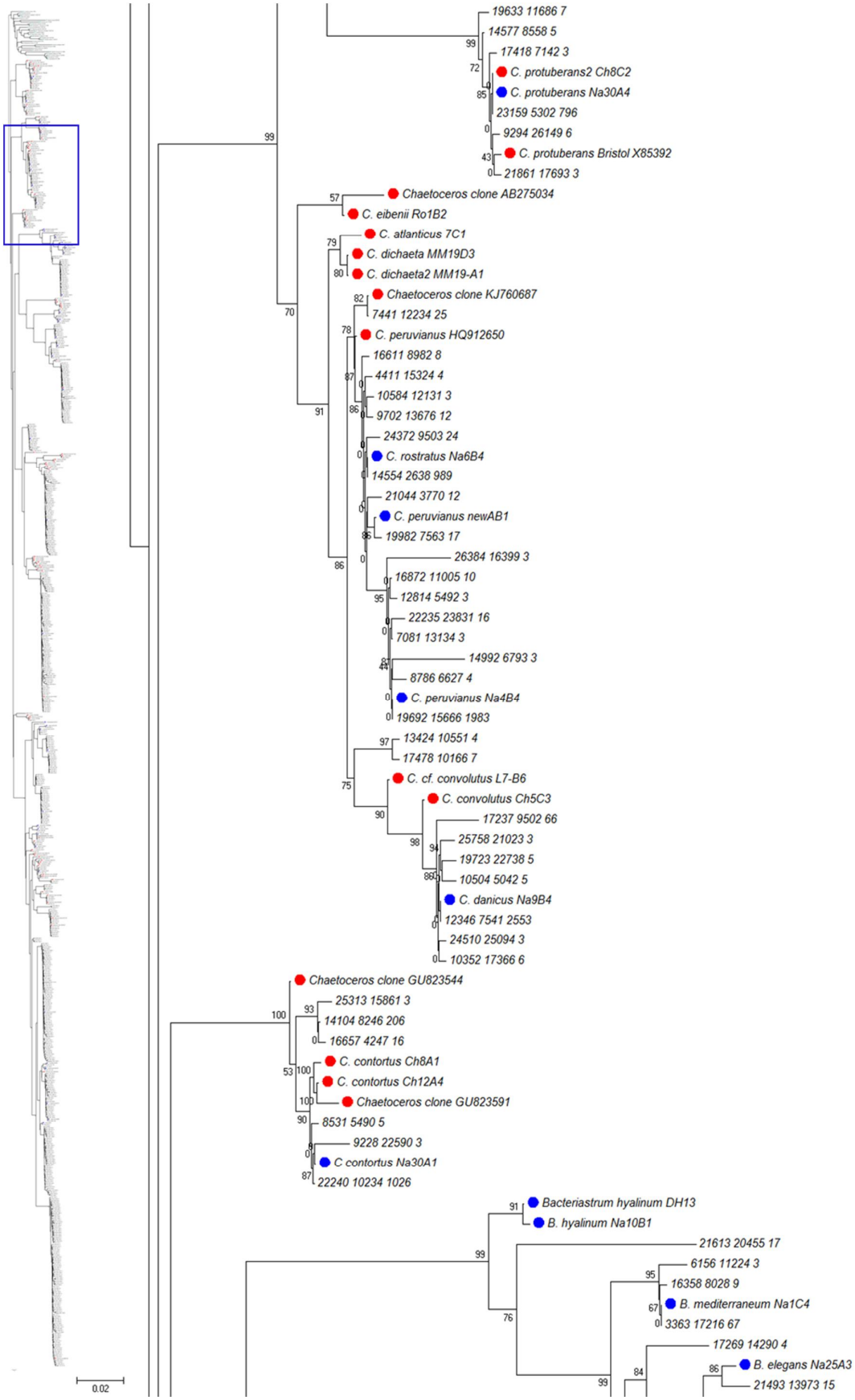
Chaetoceros curvisetus complex and sister clades

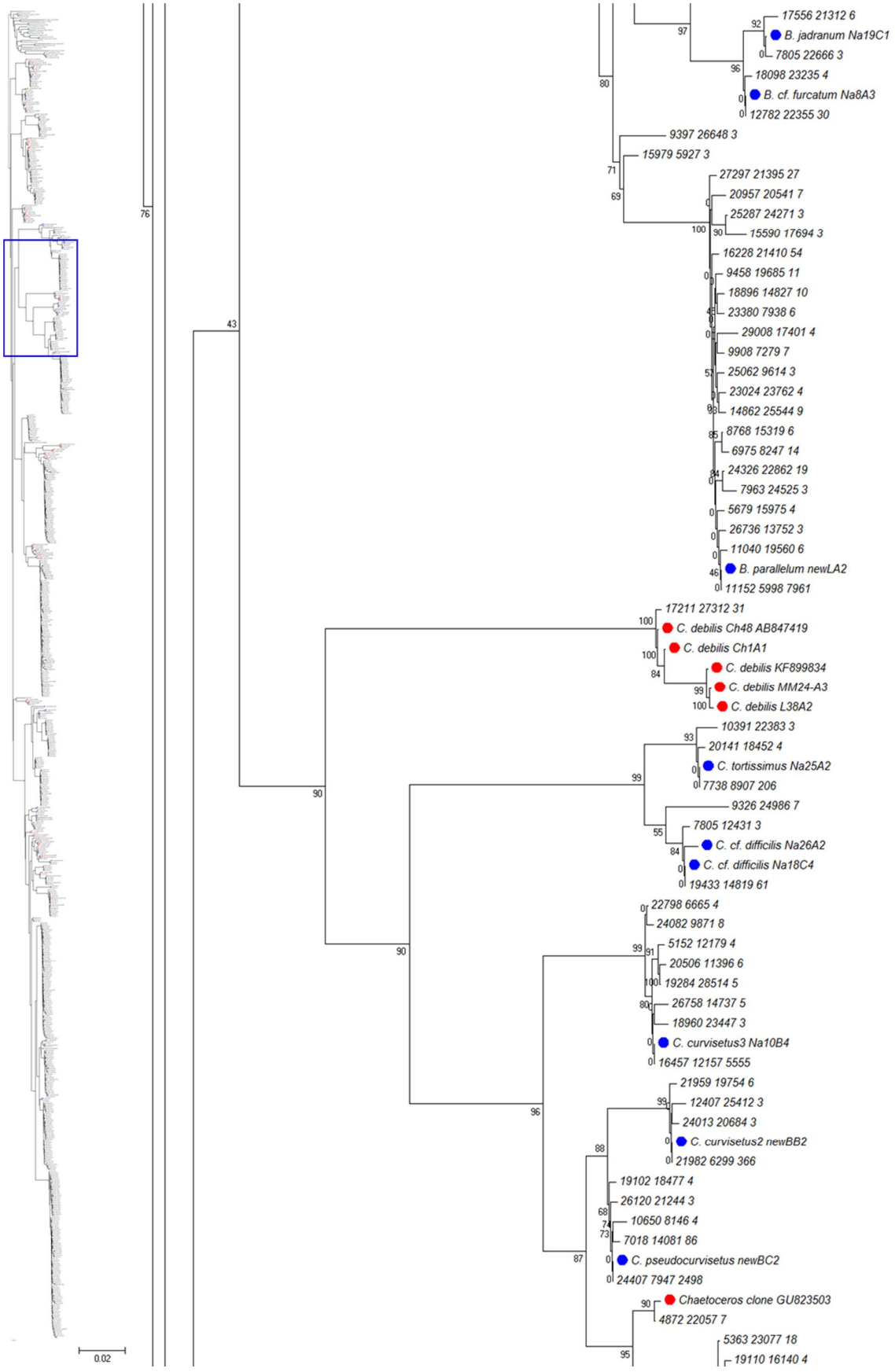
The fifth principal clade (90%) included in its turn a series of highly supported terminal clades on a well-resolved topology (87% - 100%). This clade included four distinct reference barcodes of *C. curvisetus* and one of *C. pseudocurvisetus*, each with their associated metabarcode haplotypes, grouping in a clade as sister to a clade with reference barcodes of

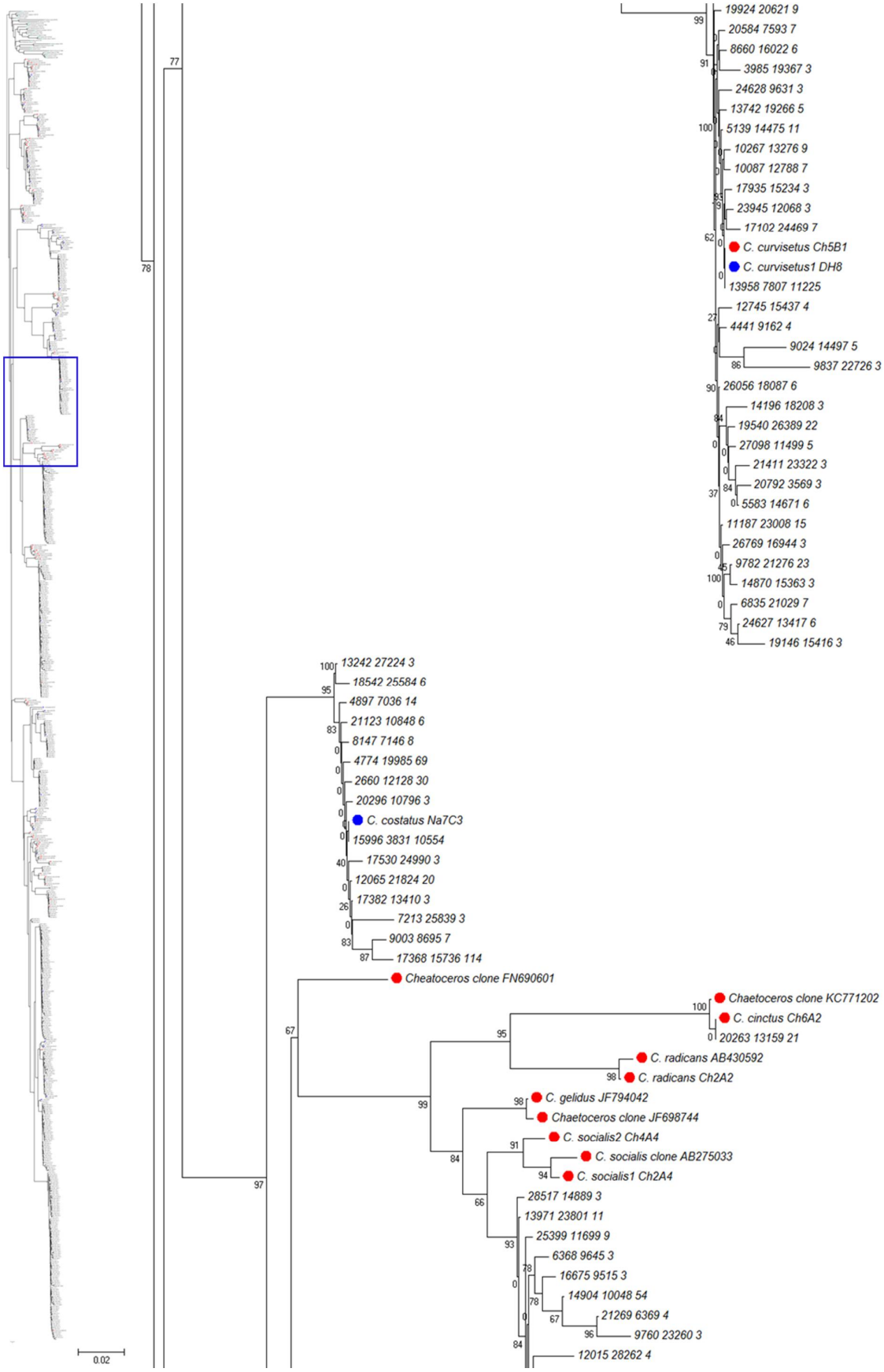
C. tortissimus and *C.* cf. *difficilis*, each with their metabarcode haplotypes. The next nearest sister has recovered a clade with reference barcodes of *C. debilis* and one metabarcode haplotype. All of these taxa belonged within the subgenus *Hyalochaete.* Within the clade of *C. debilis* -a species not known from the Gulf of Naples- a single metabarcode haplotype (*n*=31) was recovered in March and May 2012. This metabarcode haplotype matched close to a reference barcode from the Seto Inland Sea, Japan (AB847419), differing from it by only one base pair. The metabarcode haplotypes of *C. tortissimus* were encountered on and off throughout the seasons, except in winter, whereas those of *C.* cf. *difficilis* were found likewise, but throughout the year. The metabarcode haplotype, 9326_24986 (*n*=7) was recovered as a sister to the *C.* cf. *difficilis* but well resolved from the *C.* cf. *difficilis* clade. This metabarcode haplotype co-occurred with those of the abovementioned two species in May 2011, but it also was detected in Nov. and Dec. 2013 when the other two species were not detected. The metabarcode haplotypes in the four terminal clades of *C. curvisetus* and *C. pseudocurvisetus* were detected basically year-round but dominated from the autumn into the early spring. The metabarcode haplotype 4872_22057 (*n*=7) was identical to an environmental clone from the Caribbean Sea (GU823503). This metabarcode haplotype was resolved as a recombinant between the reference barcodes of *C. pseudocurvisetus* (strain newBC2) and *C. curvisetus* (strain Na1C1), its sequence aligning with the former until position 240, and with the latter thereafter. In addition, this recombinant metabarcode haplotype uniquely shared three base substitutions with the environmental clone, setting them apart from their supposed parental haplotypes.

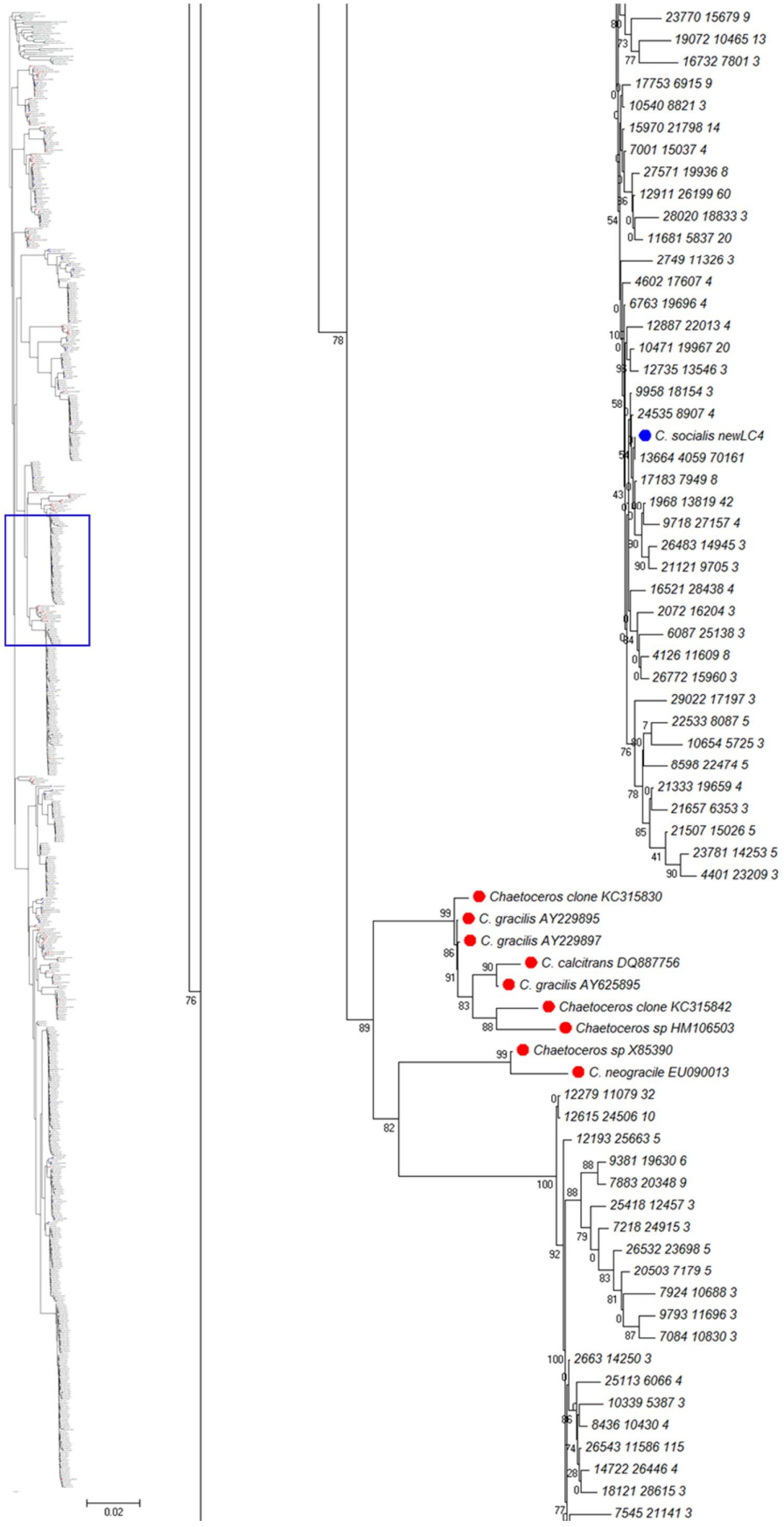
Fig. 4.2. Phylogeny inferred from the 628 Chaetocerotacean V4-metabarcode haplotypes, and 149 reference barcodes of Chaetocerotacean species and outgroups. The blue dots indicate V4 reference barcodes generated from strains originating from the Gulf of Naples, while red dots represent reference barcodes of strains and environmental sequences obtained from elsewhere. Each metabarcode haplotype is indicated with three numbers, the last of which represents the number of contigs in that haplotype.


Urosolenia eriensis Y98-8
Helicotheca tamesis
Lithodesmium undulatum CCMP1797
Bellerochea malleus
Hemiaulus sp. newAC3
Terpsinoe musica NHOP43
Hydrosera sp CYTX025
Dactyliosolen blavyanus Na14B4
Dactyliosolen cf. blavyanus ECT3891
Porosira pseudodelicatula CCMP1433
Thalassiosira gessneri ANO2-08
Thalassiosira weissflogii CCMP1010
Cymatosira belgica
Papiliocellulus simplex CS431
Mastodiscus radiatus ECT3822
Odontella aurita CCAP1054
Triceratium dubium CCMP147
Isthmia enervis CXCL
Biddulphiopsis titiana ECT3697
Toxarium undulatum WK44
Ardissonea formosa ECT3655
Attheya longicornis CCMP214
Calyptrella robusta
Trigonium formosum ECT3689
Cerataulina pelagica ECT3845
Eucampia zoodiacus CCMP386
Eucampia sp. Ch10C2
E groenlandica RCC1996
Chaetoceros clone AB827583
Chaetoceros clone AB827511
Chaetoceros clone AB827490
Chaetoceros clone GU823338
C. teres Ch6A4
C. teres Ch6B1
8563 16228 11
Chaetoceros sp. FR865486
13514 22860 4
C. lauderi Na1A1
C. lauderi Na13A4
20257 27669 456
9708 17645 14
Chaetoceros sp. ES8
25253 19609 14
12098 16403 7
27951 20433 4052
C. brevis/pseudobrevis Ch9B3
16167 25287 6
21722 6048 3
20981 15317 8
10943 23576 6
27722 16191 3
C. brevis Na3C3
10206 8263 1527
Chaetoceros sp. KJ866910
C. brevis/pseudobrevis Na7C2
24122 17797 176
11358 20938 3
15985 23628 6
Chaetoceros sp. GU823798
26115 18413 1127
C. didymus Ch2B4
8351 19007 4
5884 13610 5
C. didymus Na20B4
18656 19195 1061
19633 11686 7
0.02

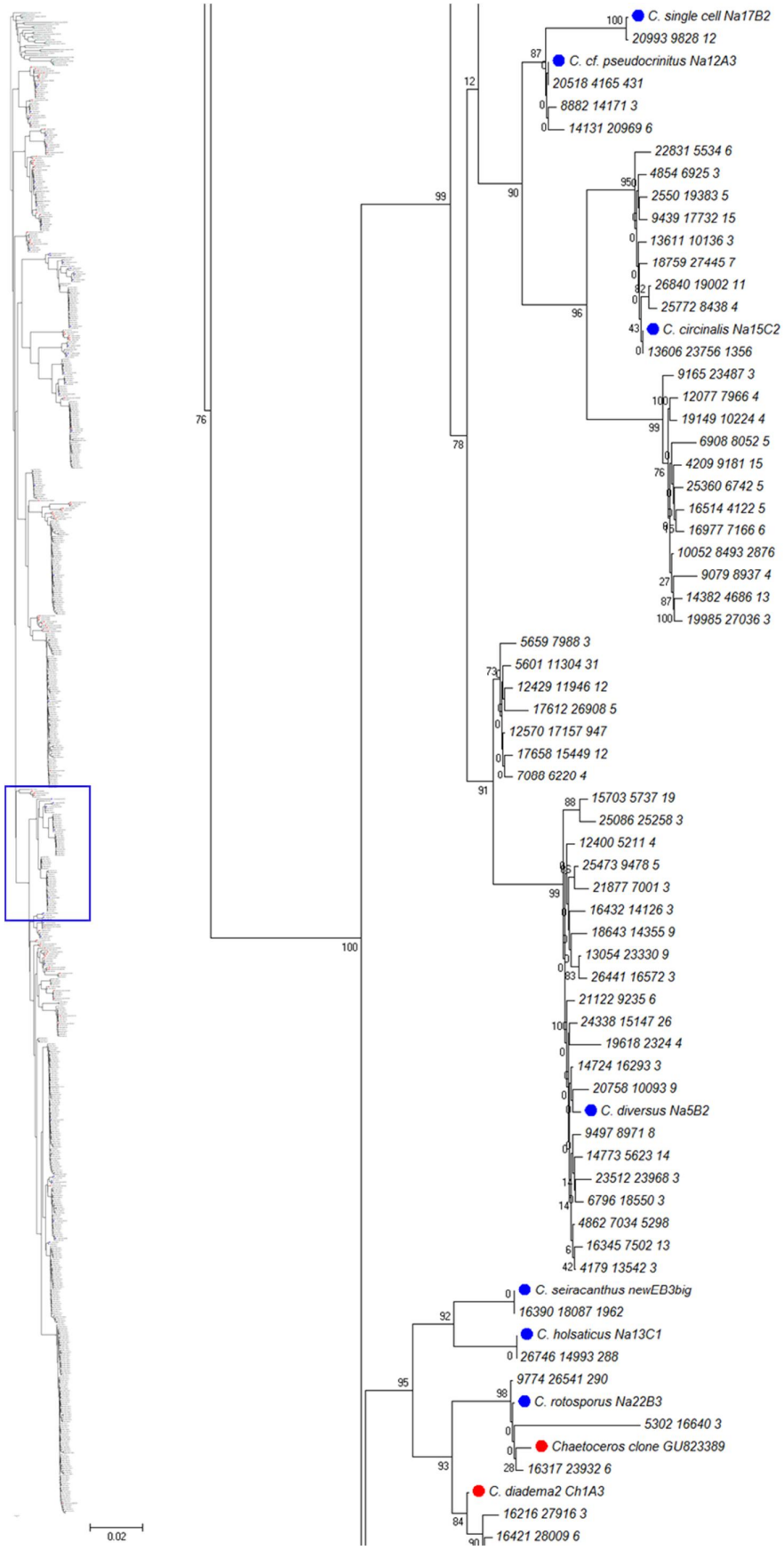

19633 11686 7
14577 8558 5
17418 7142 3
C. protuberans2 Ch8C2
C. protuberans Na30A4
23159 5302 796
9294 26149 6
C. protuberans Bristol X85392
21861 17693 3
Chaetoceros clone AB275034
C. eibenii Ro1B2
C. atlanticus 7C1
C. dichaeta MM19D3
C. dichaeta2 MM19-A1
Chaetoceros clone KJ760687
7441 12234 25
C. peruvianus HQ912650
16611 8982 8
4411 15324 4
10584 12131 3
9702 13676 12
24372 9503 24
C. rostratus Na6B4
14554 2638 989
21044 3770 12
C. peruvianus newAB1
19982 7563 17
26384 16399 3
16872 11005 10
12814 5492 3
22235 23831 16
7081 13134 3
14992 6793 3
8786 6627 4
C. peruvianus Na4B4
19692 15666 1983
13424 10551 4
17478 10166 7
C. cf. convolutus L7-B6
C. convolutus Ch5C3
17237 9502 66
25758 21023 3
19723 22738 5
10504 5042 5
C. danicus Na9B4
12346 7541 2553
24510 25094 3
10352 17366 6
Chaetoceros clone GU823544
25313 15861 3
14104 8246 206
16657 4247 16
C. contortus Ch8A1
C. contortus Ch12A4
Chaetoceros clone GU823591
8531 5490 5
9228 22590 3
C contortus Na30A1
22240 10234 1026
Bacteriastrum hyalinum DH13
B. hyalinum Na10B1
21613 20455 17
6156 11224 3
16358 8028 9
B. mediterraneum Na1C4
3363 17216 67
17269 14290 4
B. elegans Na25A3
21493 13973 15
0.02


17556 21312 6
B. jadranum Na19C1
7805 22666 3
18098 23235 4
B. cf. furcatum Na8A3
12782 22355 30
9397 26648 3
15979 5927 3
27297 21395 27
20957 20541 7
25287 24271 3
15590 17694 3
16228 21410 54
9458 19685 11
18896 14827 10
23380 7938 6
29008 17401 4
9908 7279 7
25062 9614 3
23024 23762 4
14862 25544 9
8768 15319 6
6975 8247 14
24326 22862 19
7963 24525 3
5679 15975 4
26736 13752 3
11040 19560 6
B. parallelum newLA2
11152 5998 7961
17211 27312 31
C. debilis Ch48 AB847419
C. debilis Ch1A1
C. debilis KF899834
C. debilis MM24-A3
C. debilis L38A2
10391 22383 3
20141 18452 4
C. tortissimus Na25A2
7738 8907 206
9326 24986 7
7805 12431 3
C. cf. difficilis Na26A2
C. cf. difficilis Na18C4
19433 14819 61
22798 6665 4
24082 9871 8
5152 12179 4
20506 11396 6
19284 28514 5
26758 14737 5
18960 23447 3
C. curvisetus3 Na10B4
16457 12157 5555
21959 19754 6
12407 25412 3
24013 20684 3
C. curvisetus2 newBB2
21982 6299 366
19102 18477 4
26120 21244 3
10650 8146 4
7018 14081 86
C. pseudocurvisetus newBC2
24407 7947 2498
Chaetoceros clone GU823503
4872 22057 7
5363 23077 18
19110 16140 4
0.02

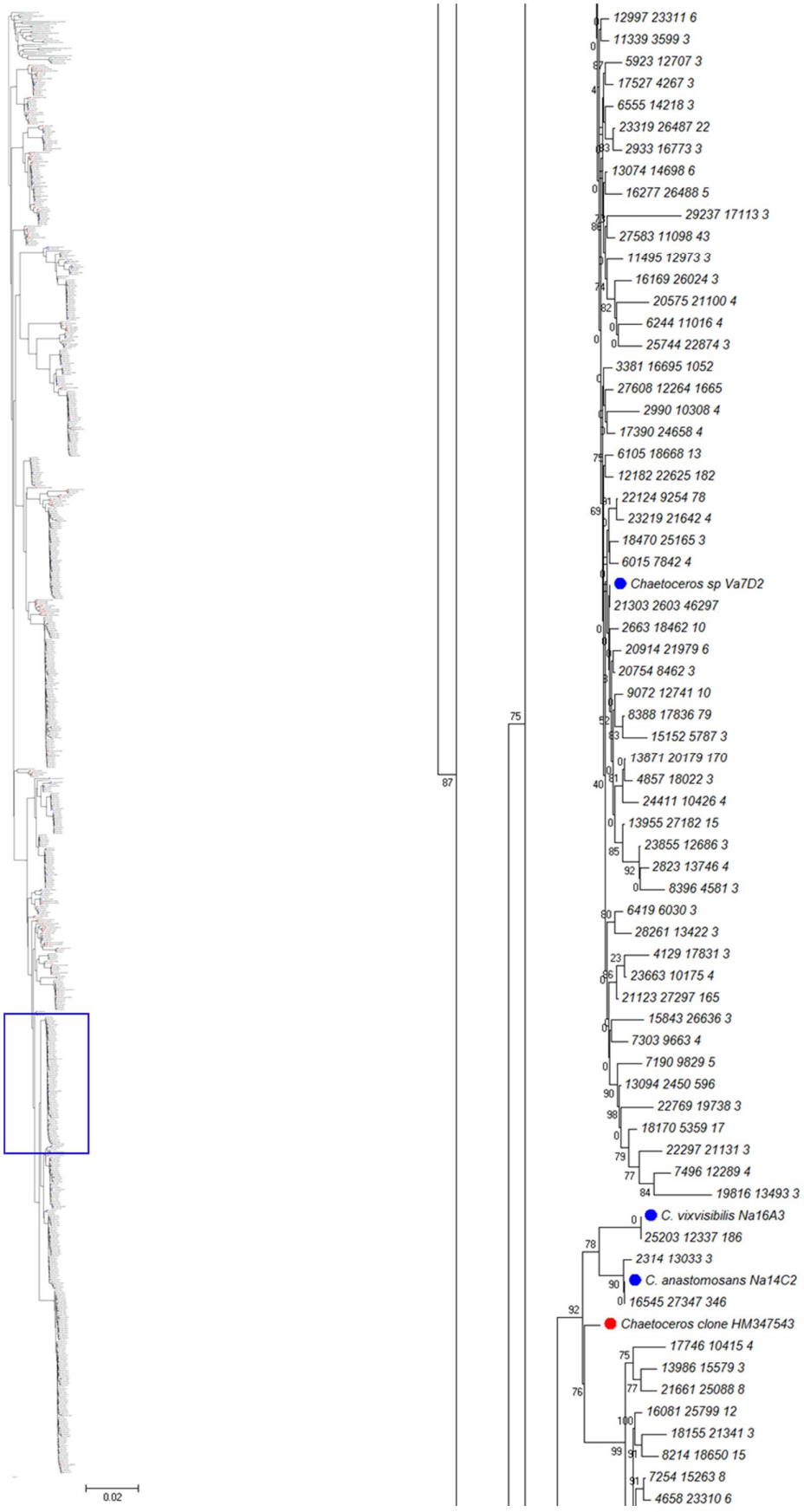

19924 20621 9
20584 7593 7
8660 16022 6
3985 19367 3
24628 9631 3
13742 19266 5
5139 14475 11
10267 13276 9
10087 12788 7
17935 15234 3
23945 12068 3
17102 24469 7
C. curvisetus Ch5B1
C. curvisetus1 DH8
13958 7807 11225
12745 15437 4
4441 9162 4
9024 14497 5
9837 22726 3
26056 18087 6
14196 18208 3
19540 26389 22
27098 11499 5
21411 23322 3
20792 3569 3
5583 14671 6
11187 23008 15
26769 16944 3
9782 21276 23
14870 15363 3
6835 21029 7
24627 13417 6
19146 15416 3
13242 27224 3
18542 25584 6
4897 7036 14
21123 10848 6
8147 7146 8
4774 19985 69
2660 12128 30
20296 10796 3
C. costatus Na7C3
15996 3831 10554
17530 24990 3
12065 21824 20
17382 13410 3
7213 25839 3
9003 8695 7
17368 15736 114
Cheatoceros clone FN690601
Chaetoceros clone KC771202
C. cinctus Ch6A2
20263 13159 21
C. radicans AB430592
C. radicans Ch2A2
C. gelidus JF794042
Chaetoceros clone JF698744
C. socialis2 Ch4A4
C. socialis clone AB275033
C. socialis1 Ch2A4
28517 14889 3
13971 23801 11
25399 11699 9
6368 9645 3
16675 9515 3
14904 10048 54
21269 6369 4
9760 23260 3
12015 28262 4
0.02

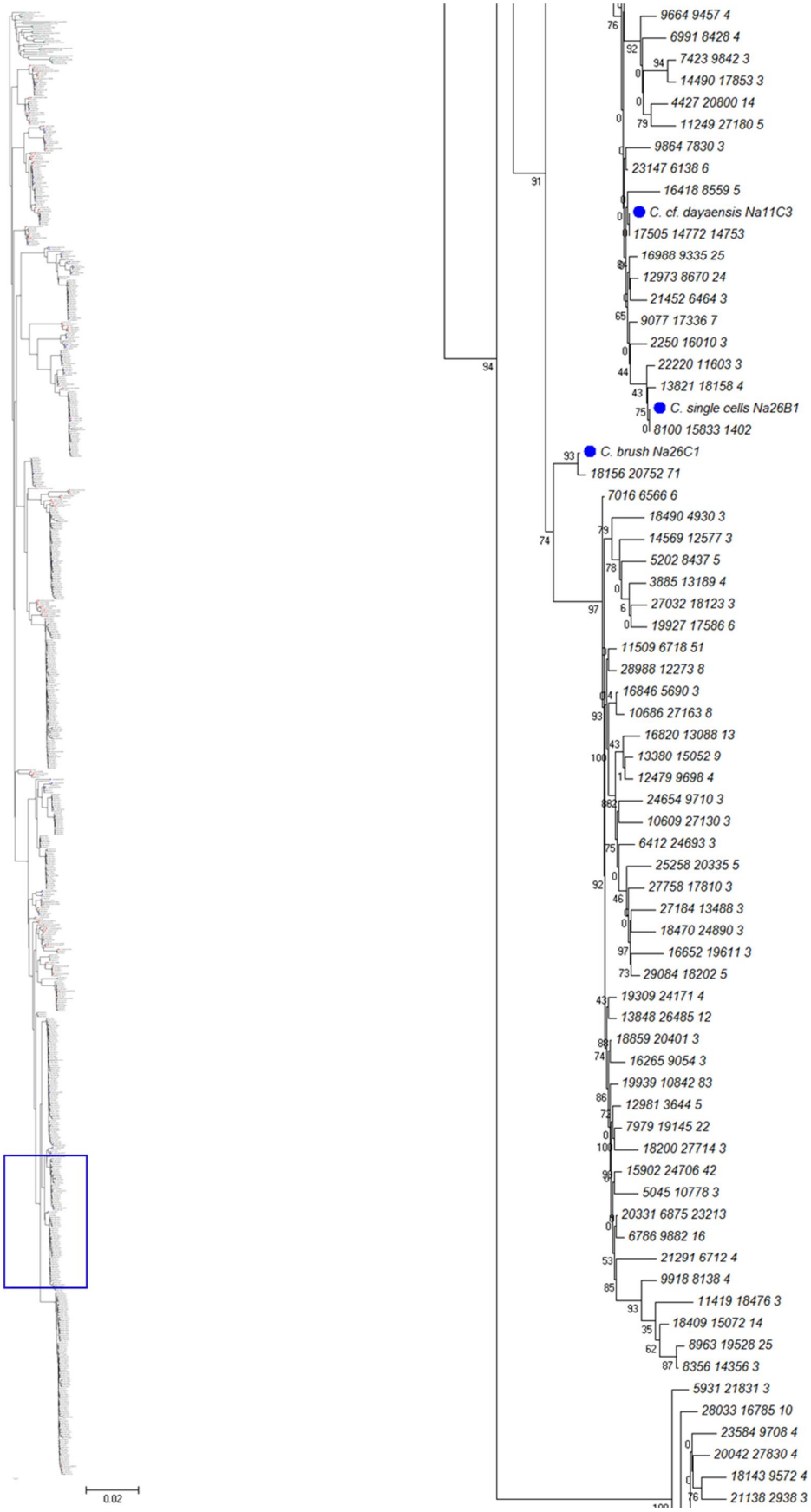

23770 15679 9
19072 10465 13
16732 7801 3
17753 6915 9
10540 8821 3
15970 21798 14
7001 15037 4
27571 19936 8
12911 26199 60
28020 18833 3
11681 5837 20
2749 11326 3
4602 17607 4
6763 19696 4
12887 22013 4
10471 19967 20
12735 13546 3
9958 18154 3
24535 8907 4
C. socialis newLC4
13664 4059 70161
17183 7949 8
1968 13819 42
9718 27157 4
26483 14945 3
21121 9705 3
16521 28438 4
2072 16204 3
6087 25138 3
4126 11609 8
26772 15960 3
29022 17197 3
22533 8087 5
10654 5725 3
8598 22474 5
21333 19659 4
21657 6353 3
21507 15026 5
23781 14253 5
4401 23209 3
Chaetoceros clone KC315830
C. gracilis AY229895
C. gracilis AY229897
C. calcitrans DQ887756
C. gracilis AY625895
Chaetoceros clone KC315842
Chaetoceros sp HM106503
Chaetoceros sp X85390
C. neogracile EU090013
12279 11079 32
12615 24506 10
12193 25663 5
9381 19630 6
7883 20348 9
25418 12457 3
7218 24915 3
26532 23698 5
20503 7179 5
7924 10688 3
9793 11696 3
7084 10830 3
2663 14250 3
25113 6066 4
10339 5387 3
8436 10430 4
26543 11586 115
14722 26446 4
18121 28615 3
7545 21141 3
0.02

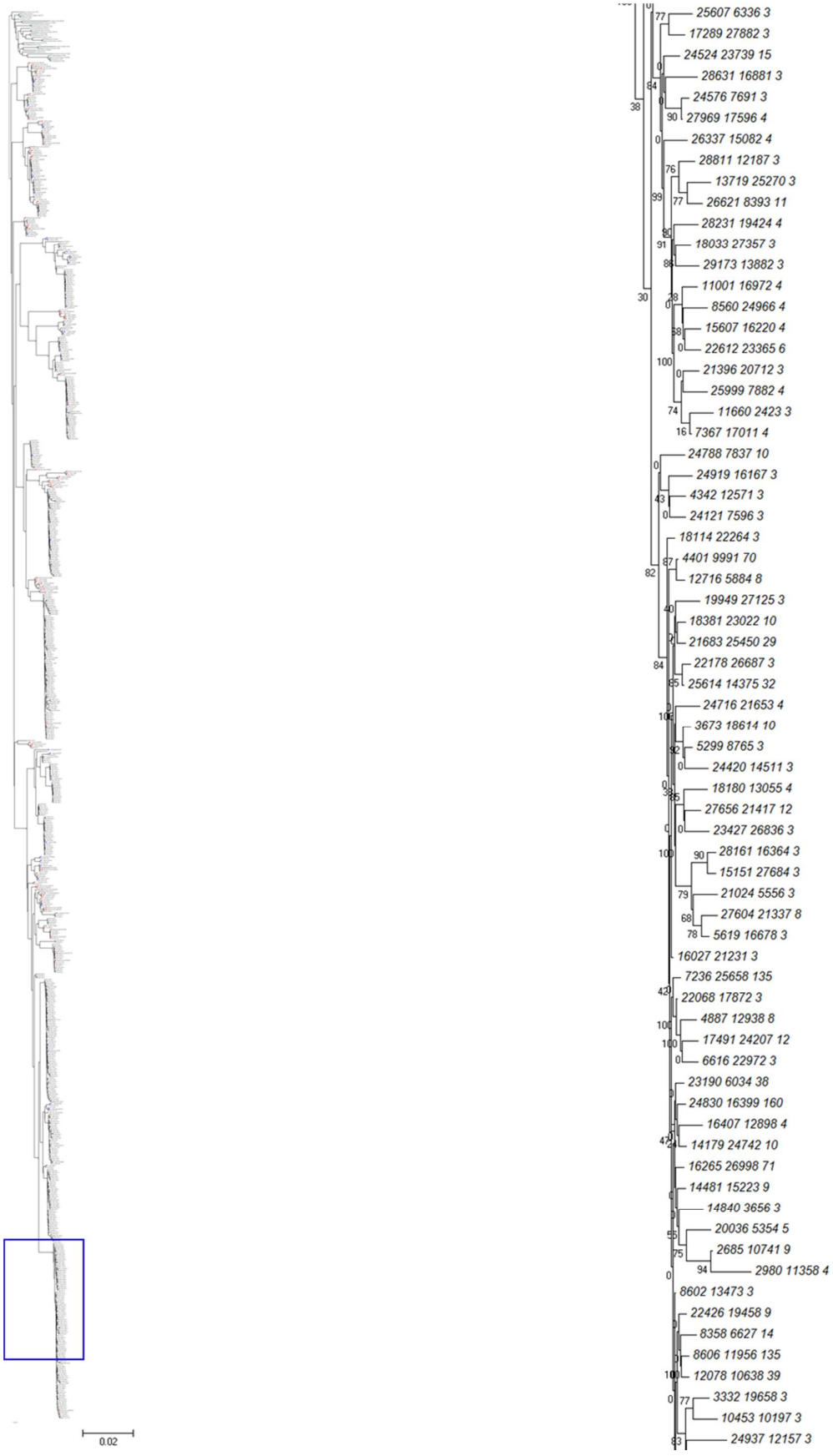

7545 21141 3
21407 8003 3
21578 12753 18
11362 11058 4
15272 7817 63
27191 20341 5
17604 20782 3
5940 19136 5
4105 19089 24
8256 20075 3
16599 9545 3
21679 26489 4
24217 11502 4
17400 5832 5
17994 25601 47
18259 8881 3
C. tenuissimus Na26A1
16520 2902 150522
5048 13343 6
29191 17492 4
11492 16939 3
10370 3752 3
2982 10183 3
12029 4420 3
18058 11717 127
9024 8158 3
3083 20673 3
5228 11036 6
6365 9495 12
20081 22806 3
6266 18230 3
9852 22271 4
18691 9136 11
13139 16806 440
26813 11962 3
20932 24586 3
21344 5111 3
27385 14305 8
5352 17962 8
19601 12507 4
25463 11005 6
6944 7259 3
2776 15723 3
23242 17832 8
11205 3688 3
13248 20412 8
16047 18965 3
26519 11992 8
20002 10813 4
6782 13775 13
19850 15560 614
8059 11160 3
Chaetoceros sp AF145226
11789 12803 6
10209 19118 3
2986 11022 5
23416 7578 50
12687 25268 13
15471 10530 6
11460 8483 19
11235 25914 3
5826 10549 5
25867 13153 3
C. jonquieri DQ830988
Chaetoceros sp EF473734
C. pumilus AY625894
24886 21051 5
C. verylongsetae Na13C2
23494 4001 4
0.02

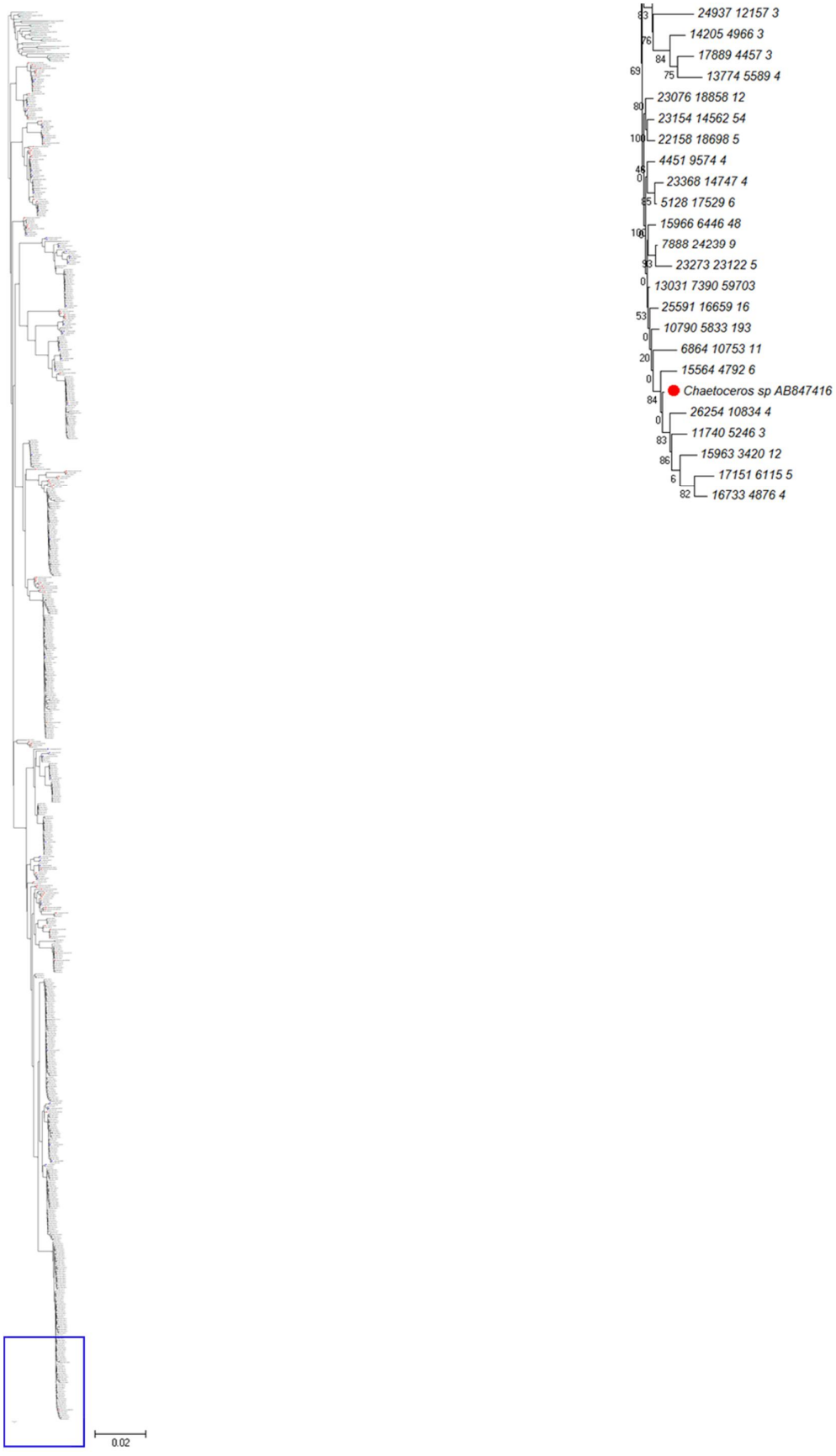

C. single cell Na17B2
20993 9828 12
C. cf. pseudocrinitus Na12A3
20518 4165 431
8882 14171 3
14131 20969 6
22831 5534 6
4854 6925 3
2550 19383 5
9439 17732 15
13611 10136 3
18759 27445 7
26840 19002 11
25772 8438 4
C. circinalis Na15C2
13606 23756 1356
9165 23487 3
12077 7966 4
19149 10224 4
6908 8052 5
4209 9181 15
25360 6742 5
16514 4122 5
16977 7166 6
10052 8493 2876
9079 8937 4
14382 4686 13
19985 27036 3
5659 7988 3
5601 11304 31
12429 11946 12
17612 26908 5
12570 17157 947
17658 15449 12
7088 6220 4
15703 5737 19
25086 25258 3
12400 5211 4
25473 9478 5
21877 7001 3
16432 14126 3
18643 14355 9
13054 23330 9
26441 16572 3
21122 9235 6
24338 15147 26
19618 2324 4
14724 16293 3
20758 10093 9
C. diversus Na5B2
9497 8971 8
14773 5623 14
23512 23968 3
6796 18550 3
4862 7034 5298
16345 7502 13
4179 13542 3
C. seiracanthus newEB3big
16390 18087 1962
C. holsaticus Na13C1
26746 14993 288
9774 26541 290
C. rotosporus Na22B3
5302 16640 3
Chaetoceros clone GU823389
16317 23932 6
C. diadema2 Ch1A3
16216 27916 3
16421 28009 6
0.02


C. diadema1 Na12C1
20662 22227 3198
C. cf. constrictus Ch12C1
17643 8947 538
Chaetoceros clone HM581779
Chaetoceros sp AB847415
Chaetoceros clone DQ310333
12946 10399 378
C. cf. lorenzianus AB847414
C. lorenzianus1 Ch4C3
C. lorenzianus1 Ch3B4
C. lorenzianus2 Ch4C4
10722 25762 3
20954 27468 3
C. decipiens Na1A1
5334 15807 1574
Chaetoceros clone GU823068
Chaetoceros clone AB827452
19937 23975 64
C. lorenzianus3 Ch12A1
11434 3757 8
6378 15573 6
20240 13059 4
13951 24548 2231
18836 9283 4
10235 21571 3
C. muellerii JF790992
11911 8280 42
Chaetoceros clone DQ310200
24123 19396 3
11590 20659 33
21917 5770 3
Chaetoceros clone EF527027
13206 2953 735
10180 19462 19
14339 9121 3
15515 11812 3
9322 26768 8
15453 21287 5
11512 23842 12
23343 10046 3
Chaetoceros clone EU371179
21928 13061 81
21267 18137 23
3156 17408 4
Chaetoceros clone FR874617
15720 12712 3125
18593 22809 102
2615 10733 12
12132 4506 3
5706 7492 4
24800 6091 3
5720 21061 35
22955 8303 5
12378 8303 19
16291 19316 3
8875 16467 5
6056 17241 3
27327 22203 4
23215 9658 5
23718 19928 3
21569 21627 15
4097 8513 4
17083 3887 9
15035 2319 3
10099 5447 4
10322 3726 6
7845 10318 28
6375 12357 3
0.02


12997 23311 6
11339 3599 3
5923 12707 3
17527 4267 3
6555 14218 3
23319 26487 22
2933 16773 3
13074 14698 6
16277 26488 5
29237 17113 3
27583 11098 43
11495 12973 3
16169 26024 3
20575 21100 4
6244 11016 4
25744 22874 3
3381 16695 1052
27608 12264 1665
2990 10308 4
17390 24658 4
6105 18668 13
12182 22625 182
22124 9254 78
23219 21642 4
18470 25165 3
6015 7842 4
Chaetoceros sp Va7D2
21303 2603 46297
2663 18462 10
20914 21979 6
20754 8462 3
9072 12741 10
8388 17836 79
15152 5787 3
13871 20179 170
4857 18022 3
24411 10426 4
13955 27182 15
23855 12686 3
2823 13746 4
8396 4581 3
6419 6030 3
28261 13422 3
4129 17831 3
23663 10175 4
21123 27297 165
15843 26636 3
7303 9663 4
7190 9829 5
13094 2450 596
22769 19738 3
18170 5359 17
22297 21131 3
7496 12289 4
19816 13493 3
C. vixvisibilis Na16A3
25203 12337 186
2314 13033 3
C. anastomosans Na14C2
16545 27347 346
Chaetoceros clone HM347543
17746 10415 4
13986 15579 3
21661 25088 8
16081 25799 12
18155 21341 3
8214 18650 15
7254 15263 8
4658 23310 6
0.02


9664 9457 4
6991 8428 4
7423 9842 3
14490 17853 3
4427 20800 14
11249 27180 5
9864 7830 3
23147 6138 6
16418 8559 5
C. cf. dayaensis Na11C3
17505 14772 14753
16988 9335 25
12973 8670 24
21452 6464 3
9077 17336 7
2250 16010 3
22220 11603 3
13821 18158 4
C. single cells Na26B1
8100 15833 1402
C. brush Na26C1
18156 20752 71
7016 6566 6
18490 4930 3
14569 12577 3
5202 8437 5
3885 13189 4
27032 18123 3
19927 17586 6
11509 6718 51
28988 12273 8
16846 5690 3
10686 27163 8
16820 13088 13
13380 15052 9
12479 9698 4
24654 9710 3
10609 27130 3
6412 24693 3
25258 20335 5
27758 17810 3
27184 13488 3
18470 24890 3
16652 19611 3
29084 18202 5
19309 24171 4
13848 26485 12
18859 20401 3
16265 9054 3
19939 10842 83
12981 3644 5
7979 19145 22
18200 27714 3
15902 24706 42
5045 10778 3
20331 6875 23213
6786 9882 16
21291 6712 4
9918 8138 4
11419 18476 3
18409 15072 14
8963 19528 25
8356 14356 3
5931 21831 3
28033 16785 10
23584 9708 4
20042 27830 4
18143 9572 4
21138 2938 3
0.02


0.02
25607 6336 3
17289 27882 3
24524 23739 15
28631 16881 3
24576 7691 3
27969 17596 4
26337 15082 4
28811 12187 3
13719 25270 3
26621 8393 11
28231 19424 4
18033 27357 3
29173 13882 3
11001 16972 4
8560 24966 4
15607 16220 4
22612 23365 6
21396 20712 3
25999 7882 4
11660 2423 3
7367 17011 4
24788 7837 10
24919 16167 3
4342 12571 3
24121 7596 3
18114 22264 3
4401 9991 70
12716 5884 8
19949 27125 3
18381 23022 10
21683 25450 29
22178 26687 3
25614 14375 32
24716 21653 4
3673 18614 10
5299 8765 3
24420 14511 3
18180 13055 4
27656 21417 12
23427 26836 3
28161 16364 3
15151 27684 3
21024 5556 3
27604 21337 8
5619 16678 3
16027 21231 3
7236 25658 135
22068 17872 3
4887 12938 8
17491 24207 12
6616 22972 3
23190 6034 38
24830 16399 160
16407 12898 4
14179 24742 10
16265 26998 71
14481 15223 9
14840 3656 3
20036 5354 5
2685 10741 9
2980 11358 4
8602 13473 3
22426 19458 9
8358 6627 14
8606 11956 135
12078 10638 39
3332 19658 3
10453 10197 3
24937 12157 3


24937 12157 3
14205 4966 3
17889 4457 3
13774 5589 4
23076 18858 12
23154 14562 54
22158 18698 5
4451 9574 4
23368 14747 4
5128 17529 6
15966 6446 48
7888 24239 9
23273 23122 5
13031 7390 59703
25591 16659 16
10790 5833 193
6864 10753 11
15564 4792 6
Chaetoceros sp AB847416
26254 10834 4
11740 5246 3
15963 3420 12
17151 6115 5
16733 4876 4
0.02

Chaetoceros costatus and sister clades

The sixth principal clade (97%) included in its turn a clade (95%) with the reference barcode of *C. costatus* and associated metabarcode haplotypes as sister to a clade, which in its turn included a series of highly supported terminal clades onto the nested set of clades (67% - 99%). Apart from *Chaetoceros* clone FN690601, this clade showed a basal dichotomy into two large clades. The first one of these large clades (99%) included reference barcodes of *C. cinctus*, *C. radicans,* and *C. gelidus* as well as several distinct ones of *C. socialis*. Those of *C. cinctus* and one of the references of *C. socialis* were accompanied by metabarcode haplotypes; the latter reference by a huge number. The metabarcode haplotypes of *C. costatus* occurred year round but were most frequent in late autumn and winter. *Chaetoceros cinctus* was never observed in the Gulf of Naples, but nonetheless, a single metabarcode haplotype (n=21) was found to occur there in Jun. and Jul. 2012-samples. The Neapolitan reference barcode of *C. socialis* grouped within a large terminal clade with numerous metabarcode haplotypes, occurring year-round. This clade was sister to a clade with reference barcodes belonging to strains identified as *C. socialis* from Sagami Bay, Japan and from Chile. The second one of these large clades (89%) also showed a well-resolved internal structure of clades, but these were mostly composed of reference barcodes without accompanying metabarcode haplotypes. Only the terminal clade (99%) with the reference barcode of *C. tenuissimus* included a large number of accompanying metabarcode haplotypes, present year round. A reference barcode of this species from Malaysia (AF145226) grouped within this terminal clade as well but differed from any of the metabarcode haplotypes.

Chaetoceros jonquieri and *C. pumilus* clade

The seventh principal clade (87%) contained reference barcodes of *C. jonquieri* and *C. pumilus* accompanied by two rare metabarcode haplotypes. One of these haplotypes,

25867_13153 (*n*=3) was only distantly related to the abovementioned reference barcodes and was observed only in the winter of 2013. The second haplotype (*n*=5) was identical to the reference of *C. pumilus* (AY625894) and was detected only in Nov. 2013.

***Chaetoceros affinis* complex and sister clades**

The eighth and final principal clade (100%) contained in its turn seven well supported secondary clades, the first of which (99%) included reference barcodes of several species within the group of *C. affinis* and *C. diversus*, and associated metabarcode haplotypes, though two of the terminal clades did not include a reference. No metabarcode haplotypes were assigned to the reference barcode of the strain Na13C2, because this barcode included an insert. A local BLAST search with the reference barcode, including its insert, as a query against the raw sequence reads, including inserts, returned 21 matches (≥98%) with raw forward and reverse sequences. The reference barcode of strain Na17B2 was identical to a metabarcode haplotype (*n*=12) found in Jun. 2012 and 2013. The metabarcode haplotypes of *C*. cf. *pseudocrinitus* and those of *C. circinalis* dominated during winter and spring. The metabarcode haplotypes within the terminal clade including the metabarcode haplotype of 10052_8493 (*n*=2,876) were detected in the summer and in early autumn, whereas those in the terminal clade including metabarcode haplotype 12570_17157 (*n*=947) were observed year round, but most frequently in spring. The Neapolitan reference barcode of *C. diversus* differed slightly from the dominant metabarcode haplotype (*n*=5,298) of this species, but nonetheless, it was recovered firmly within the terminal clade. Its haplotypes occurred from the winter into mid-summer.

The second of these secondary clades (95%) included reference barcodes of *C. diadema*, *C. holsaticus, C. rotosporus* and *C. seiracanthus*, each with their accompanying metabarcode haplotypes. The metabarcode haplotype of *C. seiracanthus* was found in late winter and early summer, that of *C. holsaticus* year-round; those of *C. rotosporus* in late summer and autumn

and those of *C. diadema* in winter.The third of these secondary clades included the reference barcode of *C*. cf. *constrictus,* isolated from Central Chile (and subsequently also from the Gulf of Naples) and its metabarcode haplotype (*n*=538), which dominated in spring. The fourth of these secondary clades (98%) included only two reference barcodes of an unknown *Chaetoceros* species.

***Chaetoceros decipiens* complex clade**

The fifth of these secondary clades (87%) included in its turn a series of terminal clades and reference barcodes of *C.* cf. *lorenzianus, C. decipiens* and a series of environmental references of unknown taxonomic provenance, all within a well-resolved topology. Metabarcode haplotype 12946_10399 (*n*=378) was found during winter and early spring. The metabarcode haplotypes of *C. decipiens* were observed throughout the year. The metabarcode haplotype coded 19937_23975 (*n*=64) grouped with an environmental sequence from the Caribbean Sea and one from the Kuroshio Current, Japan and was detected only during Sept-Oct 2013. The two rare metabarcode haplotypes grouping with the reference barcode of the Chilean *C. lorenzianus* strain Ch12A1 were observed only during the summer of 2011.

The clade with *C. decipiens* and *C. lorenzianus* was sister to a clade, which in its turn was composed of five terminal clades all with taxonomically unknown or uncertain *Chaetoceros* species. The first of these included metabarcode haplotype 13951_24548 (*n*=2,231) and occurred mainly in spring and autumn, the second included metabarcode haplotype 11911_8280 (*n*=42) and was found during the summer of 2011 and 2013. The third, which included metabarcode haplotype 13206_2953 (*n*=735), was identical or very close to two environmental clones (EF527027 and DQ310200) from Framvaren Fjord, Southern Norway. This metabarcode haplotype was detected in the Gulf of Naples from late autumn into the spring. The fourth included two metabarcode haplotypes (*n*=19) and was found only

between Jul. and Nov. of 2011. The fifth, with dominant metabarcode haplotype 15720_12712 (*n*=3,125), included environmental sequences from South-western Norway (FR874617) and from Kongsfjorden, Svalbard (EU371179). In addition, the metabarcode haplotypes in this clade were detected in the Gulf of Naples only from the winter into the early spring.

Clades with several unknown species and species new to science

The sixth of these secondary clades (83%) was composed of only three metabarcode haplotypes without a reference barcode, all of them occurring in summer.

The seventh of these secondary clades (94%) showed a well-resolved internal structure among a series of terminal clades, some of which contained very large numbers of metabarcode haplotypes, for instance, the one with the reference barcode of the enigmatic species denoted *Chaetoceros* sp. Va7D2. These haplotypes were present year round but dominated in summer. This secondary clade also included terminal clades of *C. vixvisibilis, C. anastomosans, C.* cf. *dayaensis, C.* 'brush' and two additional large terminal clades without reference barcodes. The clade of *C. vixvisibilis* was present from spring to early autumn, that of *C. anastomosans* on and off throughout the year. The clade of *C. vixvisibilis* and *C. anastomosans* resolved as sister to a large terminal clade with genetically distinct reference barcodes of two Neapolitan strains. The reference of strain Na26B1 was identical to a dominant metabarcode haplotype (*n*=1,402), and was mainly observed during summer and early autumn; whereas the reference barcode of *C.* cf. *dayaensis* strain Na11C3 was identical to another dominant haplotype (*n*=14,753) occurring year round. The terminal clade as a whole consisted of 24 peripheral haplotypes. The reference barcode of *C.* 'brush' resolved next to a metabarcode haplotype, which was found in summer and early autumn. This species was sister to a large terminal clade without any reference barcode, found likewise in summer and early autumn. The final clade in this secondary clade was also by

far the most populous one. Its dominant haplotype (n=59,703) was accompanied by a whopping 96 peripheral haplotypes, all of them occurring from late spring to early autumn. The only reference barcode grouping in this clade constituted a GenBank sequence from the Seto Inland Sea, Japan sampled in June.

The terminal clades: the Chaetocerotacean haplotypes grouped into 76 terminal clades, 66 in the *Chaetoceros* grade and ten in the clade of *Bacteriastrum*. A total of 39 of these terminal clades each included a reference barcode of a Neapolitan strain. Of these barcodes, 27 belonged to a taxonomically validated species and the remaining 12 to species new to science for which I obtained the morphological, ultrastructural and genetic information. Another seven terminal clades each included a reference barcode obtained from a cultured strain of a taxonomically validated species or a species new to science, but these strains were generated from field material collected not in the Gulf of Naples, but along the Central Chilean coast or at Roscoff (France); these taxa were never observed in the monitoring program at the LTER station MareChiara. A total of 30 terminal clades of metabarcode haplotypes lacked any taxonomically verified reference barcode, though some of these clades included one or more GenBank sequences from environmental samples.

The following general trends were observed in the composition and size of the terminal clades. Many terminal clades contained only a single metabarcode haplotype. In such cases, this haplotype represented a few to a few hundred contigs. In cases in which terminal clades included multiple metabarcode haplotypes one of these haplotypes was found to be the dominant one, representing several hundred to several hundreds of thousands of contigs whereas the other haplotypes included a far lower number of contigs (usually between 3 and a few tens). Terminal clades with dominant haplotypes containing several tens to hundreds of thousands of contigs usually included also several metabarcode haplotypes each with several tens up to several hundreds of contigs, from here on referred to as peripheral

metabarcode haplotypes. A graph (Fig. 4.3) in which the numbers of contigs in the dominant and peripheral haplotypes is plotted on a logarithmic scale as a function of the decreasing order of contig numbers of the haplotypes revealed that for each one of the clades the resulting data points grouped more or less along a straight line, except for the number of contigs in the dominant metabarcode haplotype, which was recovered up to two powers of ten above this line.

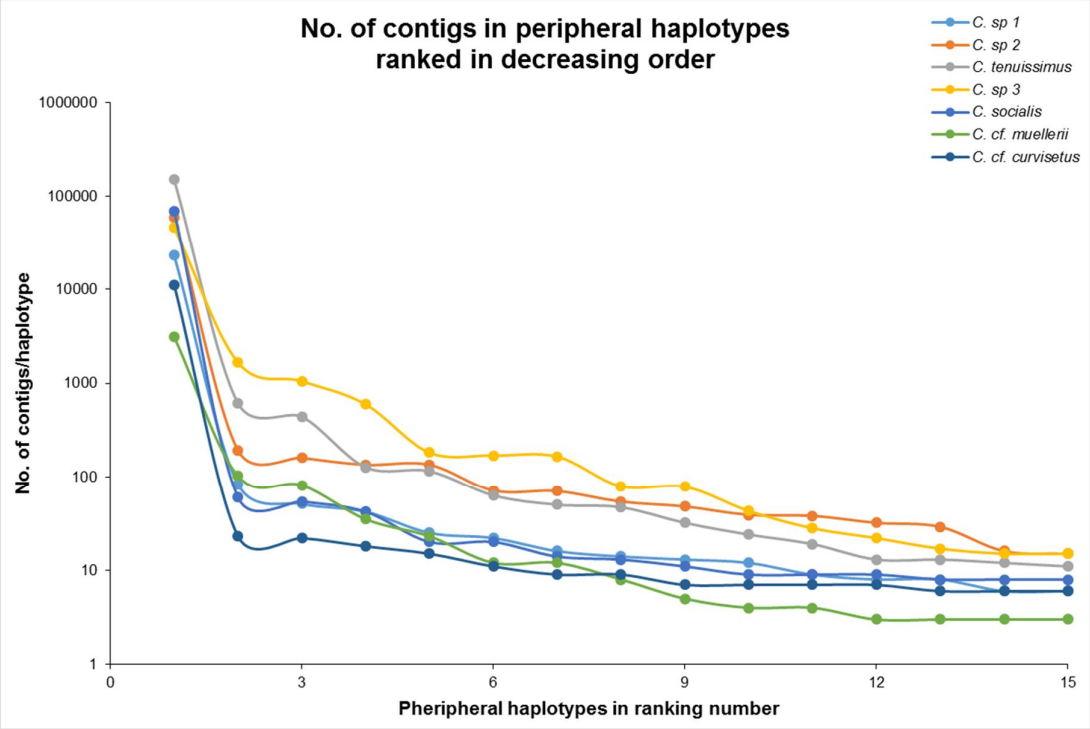


Fig. 4.3. Relationship between the numbers of contigs in the dominant and peripheral haplotypes within the seven terminal clades containing large numbers of metabarcode haplotypes. Peripheral haplotypes are characterised by containing large numbers of contigs as well. Peripheral haplotypes were ranked along the x-axis in decreasing order of contig numbers. The y-axis is scaled logarithmically.

In terminal clades containing a reference barcode of a Neapolitan strain, the dominant metabarcode haplotype was usually identical or nearly identical to this reference barcode. Reference barcodes of strains obtained from elsewhere were not necessarily identical to the dominant metabarcode haplotype; they could be recovered anywhere else in the terminal clade, i.e., differing markedly from the reference barcode.

A single terminal clade contained two genetically distinct reference barcodes; the one of strain *C.* cf. *dayaensis* Na11C3 (sampled 19/03/2014) differed in two substitutions from one of *C.* 'singlecells' Na26B1 (sampled 07/10/2014). The first of these references was identical to a metabarcode haplotype with 14,753 contigs, which occurred from Jan to Nov in all the three years, but was abundant from Jun. to Sept. The second one was identical to a metabarcode haplotype with 1,402 contigs, which occurred from the late summer into the autumn. The two strains belonged to genetically distinct sister species in the 28S rDNA tree. Similarly, the reference barcode of strains *C. rostratus* Na6B4 and *C. peruvianus* newAB1 grouped in an unresolved grade, in which only the former barcode was accompanied by a dominant metabarcode haplotype (989 contigs). Two other reference barcodes putatively assigned to *C. peruvianus* grouped in this grade as well but these originated from Texas (HQ912650; no close metabarcode haplotype) and from the East Pacific Rise (KJ760687; identical to a metabarcode haplotype with 25 contigs).

Discussion

HTS metabarcodes of the V4 region group into a large number of terminal clades with high bootstrap support and mostly well separated from other such clades. The vast majority of these clades include one or more reference barcodes (46 out of 76 terminal clades); obtained from taxonomically identified strains, and therefore, are assignable to species. The HTS-results corroborate the high Chaetocerotacean diversity in the Gulf of Naples as deduced from a light microscopic screening of the weekly samples taken at the LTER station MareChiara. They also support the strong seasonality and the recurrent annual patterns observed in the counts amassed over more than 30 years of phytoplankton monitoring (Zingone *et al.* 1995; Zingone & Sarno 2001).

Can the metabarcoding approach reveal the entire Chaetocerotacean diversity in the Gulf of Naples? Detection limits and other issues affecting species detection: species

isolated, cultured and identified during our culture-based effort over the years are generally detected also in the HTS data. Yet, exceptions abound. *Bacteriastrum hyalinum* occurs at the LTER MC because strains DH13 and Na10B1 have been collected there. In spite of that, none of the haplotypes group with these references. In addition, neither do any of the haplotypes group with the reference of *Chaetoceros* 'verylongsetae' Na13C2. One explanation for these "no-shows" in the HTS data is that species can be common in one year and exceedingly rare or even absent in another. Such patterns are discernible in the HTS data as demonstrated by, e.g., the haplotype associated to *C. brevis/pseudobrevis* Na7C2 and the haplotype close to the *Chaetoceros* clone AB827452, both of which are absent in one or two of the three years sampled for HTS. Sampling for HTS was carried out over the years 2011-2013 whereas the strains Na10B1 and Na13C2 were collected in 2014 and strain DH13 before June 2004.

Insertions in the V4 regions or in the V4-primer positions can affect species detection in HTS data. The V4 region of strain Na13C2 contains an insert of 130 bp. During the generation of the metabarcode, forward and reverse amplicons will be generated whether they contain an insert or not. However, during the processing of the amplicons into contigs, forward and reverse pairs of amplicons with an insert fail to overlap and are therefore, eliminated from the processed data. Yet, such inserts do not explain away the absence of *B. hyalinum* from the HTS data; an insert is present in the strain CCMP141 but not in the Neapolitan strains (DH13 and Na10B1). A possibility is that a population with an insert abounds in one year and another one lacking the insert in the following. Notably, all the sequenced strains of *C. diversus* (Na3C1, Na5B2, and Na15B4; all sampled in 2014 in Gulf of Naples) exhibit an insert in their V4 reference barcode, but metabarcode contigs of this species are abundant in the HTS data (sampled in 2011-2013), and these contigs are all without this insert. The fact that some of these strains showed intra-individual variation in

presence-absence of the insert across the rDNA cistrons could explain why this species showed up in the HTS data, anyway.

Primer misfit can also explain the absence of species in the HTS data. *Chaetoceros radicans* is absent in the HTS data and has never been recorded in the Gulf of Naples in the phytoplankton monitoring at the LTER-MC. The species is considered a cool-temperate species, but other such species have been observed in the Neapolitan HTS data, namely in winter and early spring. But even if the species were present, it is unlikely that it would be observed in the HTS data because of the four misfits of the reverse V4-primer with its target region, two of which are located, critically, towards the 3'-end of the primer.

The V4 region of 18S rDNA gene is highly variable and capable of amplifying taxa across eukaryotes (Hadziavdic *et al.* 2014). The V4 region is ca. 400 bp in length in most of the Chaetocerotacean species and is, therefore, suitable for sequencing with the Illumina platform. Moreover, the V4-region discriminates among all of the Chaetocerotacean taxa as reported in Chapter 3. All these taxa, which show differences in their 28S rDNA sequences, show also more than 1% sequence difference on their V4 region. However, this finding does not imply that it will tell apart all and every genetically and biologically distinct species, especially very closely related ones. The 99% similarity threshold, applied in my study pools into single haplotypes contigs that can differ up to four substitutions in the ca. 400 bp of the V4 region. Such sequence differences could be interpreted as strains belonging to different species. Since the 28S rDNA sequence region is supposed to evolve faster than the 18S V4-region, sequencing also the 28S reference barcodes of the same strains offers a way to detect such closely related species. This means that the detection limit of the marker used affects the interpretation of the results.

Terminal clades represent species: terminal clades whose dominant haplotypes include only a limited number of contigs tend to include also a small number of peripheral haplotypes

with just a few contigs, each, or none at all. Instead, terminal clades whose dominant haplotypes include huge numbers of contigs tend to be packed with large numbers of haplotypes, some of which contain numerous contigs as well. I refer to these haplotypes as peripheral ones. No fundamental difference exists between the distribution of contigs over the haplotypes in small and large terminal clades because a random sample of contigs taken from those present in a large terminal clade will predominantly include contigs from the dominant haplotype, and therefore, result in a distribution of just one or a few haplotypes, as observed within the smaller terminal clades.

The exponential decrease observed in the ranking plots of contig numbers per haplotype in all of the large terminal clades suggests that this pattern results from an equilibrium between the appearance of novel haplotypes, random drift, and the homogenizing effect of concerted evolution (Hillis & Dixon 1991; Liao 1999). The fact that the number of contigs in the dominant haplotype peaks far above the exponential trend observed in the numbers of contigs among the peripheral haplotypes probably results from the homogenizing effect of concerted evolution (Ganley & Kobayashi 2007).

Terminal clades of abundant species can thus exhibit extensive numbers of haplotypes, some of which are recovered quite distant from the dominant haplotype. These rare "outlier" haplotypes are probably just part of a "lunatic fringe." If this interpretation is correct then delineating species merely by clustering sequences in groups within a threshold perimeter of, say, 97% sequence difference from the dominant haplotype will result in several fringe-haplotypes dropping over the threshold and being counted as "rare species." Hence, the risk of seriously inflated species counts. Decelle *et al.* (2014) observed similar lunatic fringes around the dominant V4-haplotypes in individual strains of radiolarians, with several of the peripheral haplotypes dropping well below the 97% similarity threshold often used to delineate species.

Terminal clades of metabarcode haplotypes delineate species: the data confirm the expectation that a reference barcode in a terminal clade is identical to the dominant haplotype in that clade if the reference specimen is sampled from the same population from which the HTS sample has been obtained. This is expected because a Sanger-sequenced PCR-product ideally constitutes the product of all the copies in the ribosomal DNA cistrons of a specimen. If one haplotype dominates in a freely interbreeding population, as observed in the distribution of contigs over the metabarcode haplotypes in a terminal clade, then that haplotype can be expected to dominate also in all its constituent specimens.

Although contig-rich peripheral haplotypes in large terminal clades probably are a mere part of the intraspecific sequence variation, it cannot be excluded that a peripheral haplotype represents a biologically distinct species whose sequence variation is occluded-swamped by the "lunatic fringe" of its closely related species' massive terminal clade. Some of the morphologically and biologically distinct species, e.g. *C. vixvisibilis* and *C. anastomosans*, are recovered closely together. If *C. vixvisibilis* would constitute an abundant species, then the haplotypes of *C. anastomosans* might well be recovered inside the huge terminal clade of *C. vixvisibilis*, especially if ancestral haplotypes abound among the cistrons of *C. vixvisibilis*.

How to spot such occluded species? A clue constitutes the occurrence of a peripheral haplotype, offset in time from that of the dominant one. For instance, the terminal clade of *C. lauderi* includes two co-dominant haplotypes, one occurring in winter and one in summer of the years 2011-2013. Yet, all the strains I have sampled over the winters and summers of the years 2014-2015 belong to the "winter haplotype."

Another clue constitutes the presence of two genetically distinct reference barcodes, each right next to a co-dominant metabarcode haplotype in one and the same terminal clade. The prime examples constitute those of *C.* cf. *dayaensis* Na11C3 and *C. affinis*-like Na26B1,

each with their associated dominant metabarcode haplotypes. Within the 28S tree (see Chapter 3), these two strains group, indeed, in distinct clades, each of these clades containing identical sequences of additional strains. Moreover, the two groups of strains and their associated metabarcode haplotypes are sampled in different seasons. Thus, these clades represent different species.

References of the same species from elsewhere, e.g., along the central Chilean coast, often differ markedly from the dominant haplotype in the terminal clade or are recovered outside this clade. Such a distinct reference from elsewhere may represent just a different geographic population, with a different dominant haplotype, within the same biological species, or it may belong to a biologically different species. This result implies that reference barcodes from strains of local species, preferentially collected from the same samples as from which the HTS metabarcode data are generated, are a prerequisite for precise taxonomic assignment of clades in a metabarcode phylogeny. Instead, a collection of reference barcodes gathered from strains from all over the world is sufficient for the identification of terminal clades down to the level of general morphologically delineated species or closely related groups of such species. Yet, the latter will probably do not discriminate among different regional populations within a widespread meta-population.

Several terminal clades lack reference barcodes: the presence of several terminal clades without a Neapolitan reference barcode is not surprising, as several groups of specimens enumerated in the plankton counts at the LTER-MC have not yet been formally identified to the species level whereas several others that are taxonomically identified, have either not been observed or they were observed but I failed to bring them into culture.

I observed *C. dadayi, C. coarctatus* and *C. tetrastichon*, in the plankton samples in Gulf of Naples but failed to obtain cultures, and hence, reference barcodes. Nonetheless, metabarcodes of these rare species within the subgenus *Chaetoceros* can be expected to

group within the clade of this subgenus, specifically with the minor terminal clade containing haplotypes 13424_10551 and 17478_10166 and the clade with *Chaetoceros* clone KJ760687 and haplotype 7441_12234. Morphological traits of these taxa and timing of occurrence can provide predictions as to which taxon belongs to which terminal clade.

Some of the terminal clades lacking reference barcodes may represent taxonomically known species that were present in culture, but for which I failed to obtain the 18S rDNA sequences. In this case, a comparison of the topologies of the HTS-tree and the 28S rDNA tree, which includes reference sequences of additional species, allows tentative inferences about which terminal clade represents which species. For example, the terminal clade with dominant haplotype 12570_17157 is topologically in the same position as the *C. affinis* clade in the 28S-tree, the terminal clade with haplotype 10052_8493 to *C.* cf. *simplex*, and the haplotype 20331_6875 to *C. dayaensis*. These hypotheses are testable because the 18S rDNA sequences of these species can be generated and used as references.

A few large terminal clades include one or more environmental GenBank sequences or GenBank barcodes of unidentified strains. In such cases, the strains provenance and collection date are known, permitting some basic inferences about distribution patterns. For instance, I did not observe *C. throndsenii,* a species known from the LTER-MC. Nonetheless, I hypothesize that the huge clade with the reference barcode of *Chaetoceros* sp. strain SS628-11 (AB847416) collected close to the coast in the Seto Inland Sea in summer (28/06/2011) represents this species. The Seto Inland Sea is a slightly brackish estuary. At the LTER-MC the species is observed in the summer and is especially common following rainfall. Notably, the dominant metabarcode haplotype of this huge clade dominates in summer.

Another example constitutes a terminal clade with metabarcode haplotypes sampled in winter in the Gulf of Naples; it contains a reference barcode sampled in Kongsfjorden,

Svalbard (EU371179) during the Arctic summer. The occurrence of such species in Norway, or even Svalbard and the Gulf of Naples is not that strange if one considers that during midwinter (February) the surface seawater temperature of the Gulf goes down to 14 °C and the amount of sunlight reaching the water column is roughly comparable to that in the North in summer. The rest of the year such a "northern" *Chaetoceros* species can survive in the Gulf of Naples under the thermocline as resting spore in the sediment.

High Chaetocerotacean diversity in the Gulf of Naples: a reason why the Gulf of Naples harbours such a rich Chaetocerotacean flora is that it can accommodate tropical, warm temperate and cold temperate species alike, each of them encountering its favourable season and each of them overcoming the unfavourable season as resting spores. In the high summer, the stratified water column can host a diverse diatom flora thanks to nutrient-laden runoff, courtesy of a densely populated metropolitan area whereas in midwinter the water column is thoroughly mixed but sunlight is still in ample supply and the photic zone reaches all the way to the bottom, at least close to shore. Therefore, the plankton does not "rest" in the winter season in the Gulf of Naples (Zingone *et al.* 2010).

A surprising outcome in the HTS data is the detection of species known from elsewhere but not known to occur at the LTER-MC, such as *C. cinctus, C. constrictus, C. debilis* and *C. rotosporus*. Some of these species can easily be overlooked during routine monitoring because of their apparent rareness and/or resemblance to common ones already known from the Gulf of Naples. The detection of several clades in the *C. diadema*-clade in the HTS data prompted me to focus on isolating many additional strains belonging to this group.

Nonetheless, not everything is in the Gulf of Naples! According to the HTS data, numerous *Chaetoceros* species collected elsewhere appear absent from the Gulf of Naples: for e.g., *C. eibenii, C. atlanticus, C. dichaeta,* and *C. gelidus.* Several *Chaetoceros* species or species complexes exhibit references of isolates from elsewhere, resolving firmly outside the

terminal clade of Neapolitan haplotypes (as in *C. contortus, C. lorenzianus*-complex, *C*. cf. *socialis*). Whether these foreign references represent different cryptic species or mere geographic variation within widely distributed species remains to be elucidated with other approaches, but in any case, the data show that these geographic isolates do not share the same gene pool.

According to the latest update by Algaebase, ca. 230 (Guiry & Guiry 2016) the family Chaetocerotaceae contains ca. 230 taxonomically validated species globally, whereas the reference dataset contains only 80 species from the GoN (40 alone in GoN), central Chile and Roscoff: while the HTS-tree revealed 76 terminal clades in the GoN, among which many are new to science. In any case, not all these 230 taxonomically validated species can be present in the Gulf of Naples because they do not fit the metabarcode data. These taxa could be hidden in the rare contig diversity, but a preliminary check of a subsample of the ca. 18,000 haplotypes with one or two contigs showed that these contigs all grouped inside existing clades.

The number of species in Chaetocerotaceae: diatoms are arguably among the most diverse protists (Round *et al.* 1990; Guiry 2012). Estimates of their species numbers differ markedly depending on how species are defined. Guiry (2012) estimates 20,000 species with ca. 12,000 taxonomically described ones and 8,000 still to be discovered, whereas according to Mann & Droop (1996) there are likely to be over $2x10^5$ based on a fine-grained taxonomy taking into account cryptic and geographic species (Vanormelingen *et al.* 2008).

Our results support the existence of cryptic diversity, for instance in *C. curvisetus*, but this diversity was also discovered by the culture-based approach. In addition, the presence of multiple species in the *C. lorenzianus*-complex, and the *C. diadema*-complex were already noted. If only local diversity is taken into account, the number of species detected with the

HTS approach is roughly twice that with morphology. In other words, this raises the number of species with a factor 2 (as stated by Guiry 2012).

However, below the level of morphologically distinct species, my HTS data confirm evidence for the existence of fine-grained geographical and temporal variation in common species, for example in the *C. contortus*-complex as shown by Chamnansinp *et al.* (2015). Similar variation seems to exist in *C. lauderi*. If this kind of geographic patterning is common among all of the taxonomically recognized morphologically defined Chaetocerotacean species, then the number of actual species in the family is easily a factor 10 higher than what has been described. If extrapolated to the diatoms as a whole then $2x10^5$ may not be far off the mark (Mann & Droop 1996; Vanormelingen *et al.* 2008).

Abundance of species in cell counts and HTS-results: I did not perform a detailed comparison of species abundances counted in LM (Diana Sarno, Isabella Percopo) and their contigs in the HTS data. In the latter, the abundance of a species in different samples is not linearly related with the numbers of contigs assigned to that species in the HTS-results of these samples. Two samples with the same number of specimens of that species may produce different numbers of contigs depending upon the number of rDNA copies of the entire eukaryotic diversity in the sample competing for inclusion in the HTS run of that sample. Given equal numbers of a species in two samples, the total amount of DNA in the sample is negatively correlated with the number of amplified copies of the species. Also, contig numbers among different species within the same sample cannot be compared directly because of different rDNA copy numbers in the genomes of different taxa (Prokopowich *et al.* 2003). This is because of correlation between body size and rDNA copy numbers of these genes compared to others, especially in dinoflagellates and diatoms (Prokopowich *et al.* 2003; Zhu *et al.* 2005; Godhe *et al.* 2008; de Vargas *et al.* 2015).

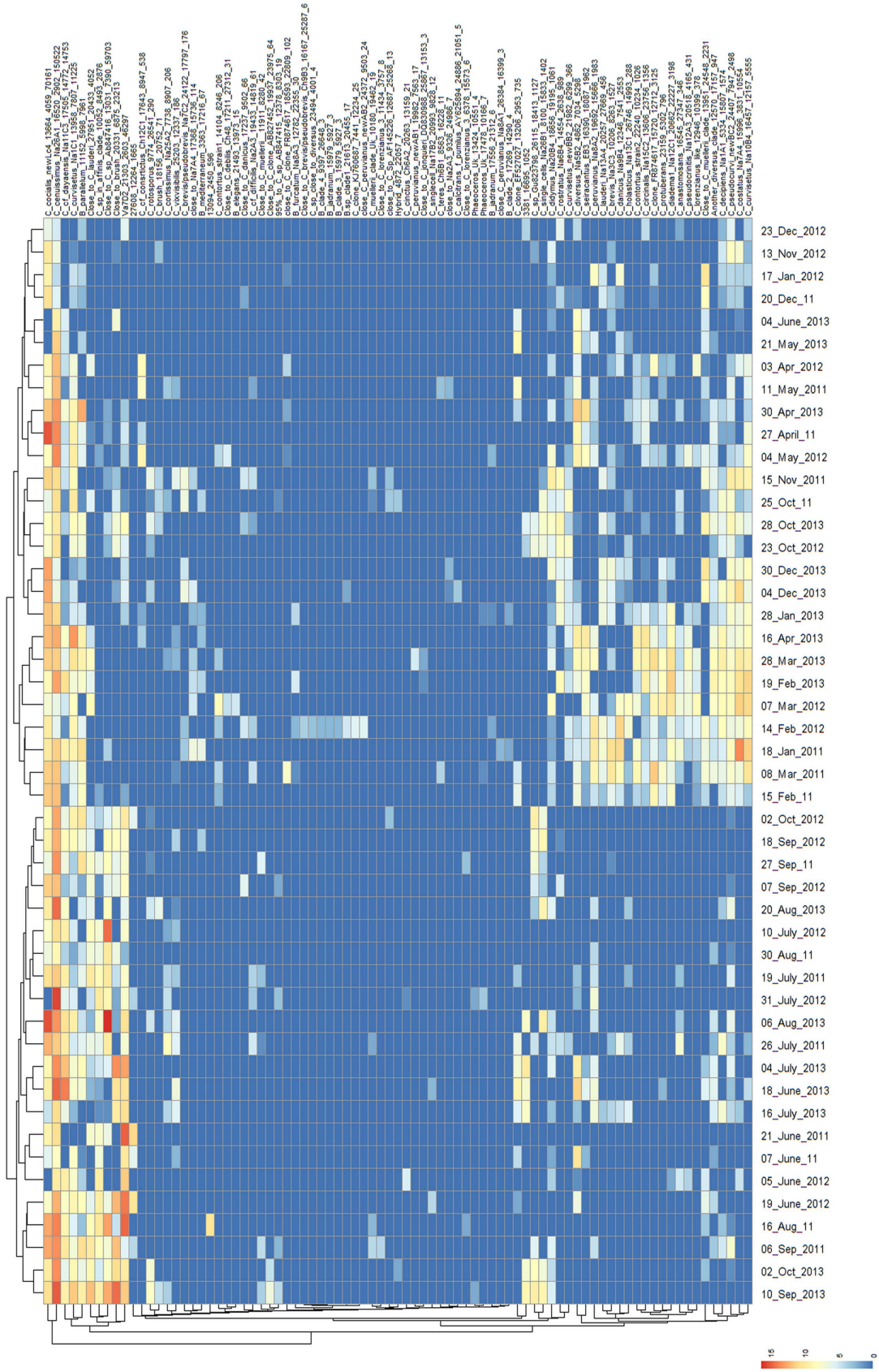
Conclusions and outlook

The comparison between culture-based- and HTS metabarcoding approaches has shown that new species can be detected with either one, but that the two approaches combined allow targeted detection of the species still to be discovered and described. The HTS results reveal when unknown taxa can be found. Once the period in which such an "unknown" occurs, one can look for the morphology of possible candidate species in fixed samples taken at the same date as the HTS sample was collected. In the following year, one can spot these taxa and characterize them by deploying the culture-based approach. Pánek *et al.* (2015) applied this combined approach successfully to uncover a high diversity of jakobids in anoxic habitats. A surprising finding was that the unknowns included extremely common *Chaetoceros* species at the LTER, and it means that we have to use the culture-based approach to properly characterize these unknowns. The HTS method can thus give semi-quantitative insight in all of the Chaetocerotacean diversity present at a site. With the costs of HTS coming down, HTS metabarcoding may replace classical monitoring by microscope screening, but only if the local diversity is properly referenced (Visco *et al.* 2015).

The present results have demonstrated that species cannot be identified securely with just a single short barcode sequence. A local population of a species exhibits a dominant haplotype, usually identical to the barcode reference obtained by means of Sanger sequencing, and a swarm of peripheral ones, some of these being quite common, but most of them exceedingly rare. Results of metabarcoding studies conducted elsewhere may provide insights how such populations of cistrons differ among geographic populations of a species. Another aspect that can be studied is how the diversity distribution of contigs over the various haplotypes in multicopy genes under the influence of concerted evolution differs from such populations of contigs over haplotypes in single copy genes, which obviously are not affected by concerted evolution.

The results also hint at why the use of short barcode probes on microarrays may fail to discriminate among closely related species such as *Pseudo-nitzschia*. The results obtained in the present study show that a reference barcode from a variable part of an SSU sequence obtained with Sanger sequencing is identical to the dominant haplotype present in a population. However, there are also several less abundant haplotypes up to a few base changes away from this dominant haplotype, some of which might be ancestral haplotypes shared with sister species. Moreover, the results demonstrate that reference barcodes sampled from different geographic strains within a single perceived species can differ from each other. If then the probes are unluckily chosen, they may fit not only the target in specimens within the local population of the intended species but also minor haplotypes in related species, or they may not fit the slightly deviating target in specimens belonging within different geographic or even different seasonal populations of the same species.

Fig. 4.4. Heatmaps showing the seasonal patterns and inter-annual changes in the abundance of the different Chaetocerotacean groups considered in Gulf of Naples, Italy.


C_socialis_nevUcd_13664_4959_70161
C_radicans_Na26A1_16220_2902_150522
C_cf_dayaensis_Na11C3_17505_14772_14753
C_curvisetus_Na1C1_13958_7807_11225
C_paradelium_11152_5980_7661
B_sp_clade1_Na1B2_20433_4052
C_sp_n_affinis_clade_10052_8493_2876
Close_to_C_sp_Ab847416_13031_7390_59703
Close_to_brush_20131_6875_23213
Yantai_2013_Na2A_16297_4289
27608_12264_1685
C_cf_constrictus_Ch12C1_17643_8947_538
C_rotosporus_s_27726541_290
C_radicans_Na1C2_7238_8907_206
C_constrictus_Na25A2_7738_8907_206
C_vixvisibilis_25203_12337_186
C_cortisimus_Na25A2_7738_8907_206
C_brevisjseudobrevis_Na17C2_24122_17797_176
C_brevis_Na1A4_17585_7525_114
B_mediterraneum_3363_17216_37
13094_2450_596
C_contortus_strain_1_14104_6246_206
Close_to_C_brevis_Ch1A1_17411_27313_31
B_elegans_21493_13975_15
Close_to_C_danicus_17237_9502_86
C_cf_difficilis_Na2A3_19433_14819_61
Close_to_C_danicus_11151_6080_12
Close_to_C_clone_AB827452_19937_23975_64
95%_to_C_sp_AB847415_12378_8303_19
close_to_C_clone_FR874617_15533_22809_102
Close_to_C_brevis_Na1A4_17585_7525_114
Close_to_C_brevis/pseudobrevis_Ch9B3_16167_25287_6
C_sp_close_to_diversus_23494_4001_4
B_clade_4_5397_8648_3
B_clade_2_15978_5927_3
B_clade_3_15979_5927_5
B_sp_clade1_21613_20455_17
C_clone_FJ760687_7441_12234_25
close_to_C_sp_AF142226_12687_25268_13
C_muelleri_clade_UK_10180_19462_19
Close_to_C_sp_AF142226_12687_25268_13
close_to_C_sp_AF142227
C_cinctus_Ch6A2_20283_13159_21
C_peruvianus_nevAB1_19982_7683_17
Close_to_C_pinquen_DQ830868_25687_13153_3
close_to_C_sp_AB827452_24122_9503_24
C_teres_Ch6B1_8563_16228_11
close_to_Na2A3_9326_24986_17
C_calcitrans_f_pumilus_A1625894_24886_21051_5
C_calcitrans_f_pumilus_A1625894_5573_6
Phaeocerros_UK_13424_10551_4
Phaeocerros_UK_17478_10186_7
Phaeocerros_UK_17356_21312_9
B_adranum_17356_21312_9
B_clade_2_17289_14290_4
C_clone_EF527027_13206_2953_735
3381_16585_1052
C_sp_Ab847415_26115_18413_1127
C_single_cells_Na28B1_3100_15833_1402
C_didymus_Na20B4_16656_19195_1061
C_rostratus_Na6B4_14554_26382_889
C_curvisetus_Na1C1_13958_7807_11225
C_diversus_Na582_4862_7034_5298
C_seiracantus_EB3_16390_18087_1962
C_peruvianus_Na8A2_19982_7683_17
C_brevis_Na3C3_10206_6263_1527
C_danicus_Na9B4_12346_7541_2553
C_holsaticus_Na13C1_26746_14993_1288
C_cf_wighamii_Na2A1_17526_7238_1076
C_circinalis_Na15C2_13696_23756_1356
C_clone_FR874617_15720_12712_3125
C_protuberans_23159_5302_796
C_cf_dayaensis_Na11C3_17505_14772_14753
C_anastomosans_18545_27547_346
C_pseudocrinitis_Na12A3_20516_4165_431
C_curvisetus_like_12946_10398_376
C_cf_wighamii_Na2A1_17526_7238_2548_2231
Another_diversus_clade_12570_17157_947
C_decipiens_Na1A1_5334_15807_1574
C_pseudocurvisetus_nevBC2_24407_7947_2498
C_cf_wighamii_Na2A1_17526_7238_1054
C_curvisetus_Na10B4_16457_12157_5555
23_Dec_2012
13_Nov_2012
17_Jan_2012
20_Dec_11
04_June_2013
21_May_2013
03_Apr_2012
11_May_2011
30_Apr_2013
27_April_11
04_May_2012
15_Nov_2011
25_Oct_11
28_Oct_2013
23_Oct_2012
30_Dec_2013
04_Dec_2013
28_Jan_2013
16_Apr_2013
28_Mar_2013
19_Feb_2013
07_Mar_2012
14_Feb_2012
18_Jan_2011
08_Mar_2011
15_Feb_11
02_Oct_2012
18_Sep_2012
27_Sep_11
07_Sep_2012
20_Aug_2013
10_July_2012
30_Aug_11
19_July_2011
31_July_2012
06_Aug_2013
26_July_2011
04_July_2013
18_June_2013
16_July_2013
21_June_2011
07_June_11
05_June_2012
19_June_2012
16_Aug_11
06_Sep_2011
02_Oct_2013
10_Sep_2013
15
10
5
0

Morphological and molecular diversity of Chaetoceros socialis Lauder

Abstract

Chaetoceros socialis is one of the most common diatoms in the plankton of the Gulf of Naples, Italy (GoN) throughout the year. It is easily recognised because of its morphological characteristics; curved chains of cells often form globular colonies. Of the four setae of each cell, three are short and curve towards the colony exterior, whereas the remaining one, which is markedly longer, protrudes towards the colony centre and joins with other long setae of adjacent cells. The main aim of the present study was to investigate whether this morphologically defined species is genetically and ultrastructurally a single species or a species complex and to reveal importance of spore morphology in species identification and discrimination. In order to address this question, several strains belonging to this entity were collected over different seasons in the GoN as well as from geographically distant places (Las Cruces and Concepcion, central Chile and Roscoff, France). All the strains were characterised morphologically and genetically. From the molecular information, it appears that all the strains collected from the GoN all over the seasons were genetically identical, and the strains collected from Roscoff, France grouped with them. Instead, the strains collected from central Chile differed genetically from the Neapolitan ones, and moreover, grouped into two sister clades, thus indicating the existence of a species complex. These two groups of strains are identified as belonging to two species new to science: *C. sporotruncatus* sp. nov. and *C. dichatoensis* sp. nov. They are morphologically similar to *C. gelidus* and *C. socialis sensu stricto*. The only morphological characters that can be used to differentiate the species of this complex are aspects of resting spore morphology. This study also characterized *C. radicans* and *C. cinctus,* which are sister species to the *C. socialis* complex. The description of *C. cintus* is vague in literature, and therefore an amended morphological description of this species was incorporated in this study.

Introduction

Chaetoceros socialis is one of the most common diatoms in the marine plankton (e.g., Sancetta & Calvert 1988; Degerlund & Eilertsen 2010; Zingone *et al.* 2010; Piredda *et al.* 2017) from polar to tropical waters (Ostenfeld 1913, Rines & Hargraves 1988). As in most *Chaetoceros* species, each cell of *C. socialis* exhibits four setae, but its characteristic feature is that three of these setae are relatively short whereas one is much longer (see Fig. 5.1 a) or sometimes all setae are short (Fig. 5.1 b). Cells form curved chains (also referred to as primary colony), and the long seta of each cell points towards the centre of the curve, whereas all the others point outwards. Chains often remain together and are connected with the long setae, thus forming a secondary spherical structure (Figs 5.1 c, d). Each cell possesses a single large plastid. Resting spores are known for this species.

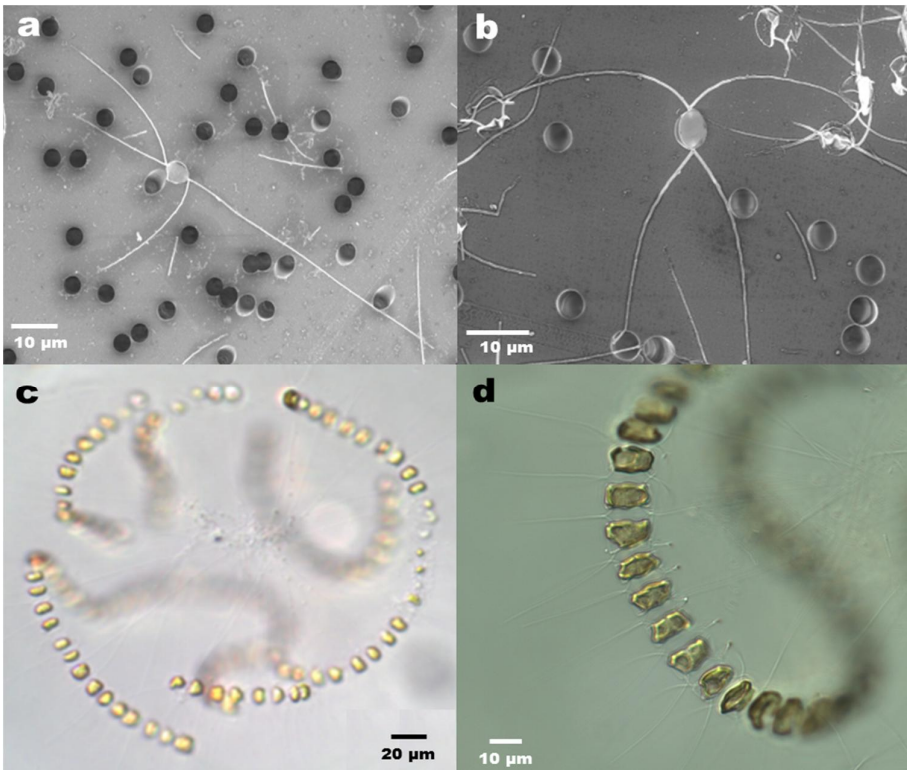


Fig. 5.1. Morphological characteristics of vegetative cells of the *Chaetoceros socialis* complex. *Chaetoceros sporotruncatus* sp. nov., strain Ch9C4 from Concepción, Chile. (a-b) SEM and (c) LM images. (a) Intercalary cell with three short setae and one long seta. (b) Intercalary cell with short setae. (c) Spherical colony; the long setae converge towards the centre of the colony. (d) A part of the spherical colony.

Chaetoceros socialis was described by Henry Scott Lauder in 1864 (Lauder 1864) from the warm waters of Hong Kong, during his tenure as an Assistant Surgeon in the Royal Navy. His description of the vegetative cells was based solely on light microscopy:

> "*Chaetoceros socialis*, n. sp. Filaments slender, aggregated, embedded in gelatine, with wavy, spirally dotted awns, some of which are more elongated, and converge to a common centre. This is the smallest species I have seen. By the aggregation of the filaments in gelatine, it forms roundish, flattened fronds. Frustules quadrate, with an awn from a little within each angle, one of them being more elongated, varying in length, according to the distance of the frustules, to a common centre, to which the elongated awns converge; many frustules, however, occur in which the awns are not thus connected; side view oval".

Lauder's description of *C. socialis* lacked information on cell dimensions.

A similar species, *C. radians* (a.k.a *C. socialis* var *radians*), was described by Schütt from material collected in the Baltic Sea. However, Schütt did not compare this species with the description of *C. socialis* from Lauder. *Chaetoceros radians* was described with the setae emerging from inside of the valve margins and with high apertures with a shallow central invagination. Cells were slightly wider than long in girdle view, with right-angled corners. Schütt also provided morphometric measurements, i.e. width 10 µm; width: length ratio 1:0.8. The seta bases were said to be curved and directed outwards, with a filamentous distal part. Cells possessed a single plate-like chloroplast. Resting spores were short, approximately spherical, with their valves covered with spines (Schütt 1895).

Cleve (1896) analysed material collected from the Baffin Bay and the Davis Strait in the Canadian Arctic and reported specimens similar to *C. socialis* but with narrow apertures. Other morphological characters of the vegetative cells did not reveal any significant

difference from *C. socialis*. However, the major difference found by Cleve was the presence of smooth resting spores in the specimens from the Arctic. Cleve stated that he was not sure about the identity of his material because Lauder's description of *C. socialis* did not include information on resting spore morphology. However, Cleve tentatively identified his material as *C. socialis*.

Following Cleve's investigation, Gran (1897) reported *C. socialis* from the west coast of Norway and *C. radians* from Drøbak. He mentioned a few differences in his finding of *C. socialis,* including valve shapes i.e., flat valves or with slight convexity in the middle with narrow aperture. Resting spores were reported as "smooth, valves almost alike, and rounded conical in shape or in a hump". For *C. radians*, he mentioned only the differences, i.e. the "wider aperture - almost as wide as the cell - and in the spores, which, in the primary valve, carry short spines". He also stated that *C. socialis* was a cosmopolitan species, having been recorded in Hong Kong, the Atlantic Ocean, and the Arctic Ocean. Ostenfeld (1913) reported *C. socialis* as an Arctic species but occurring during spring in temperate areas. He found smooth resting spores during spring and spiny resting spores during autumn in the temperate waters. For many years, *C. radians* and *C. socialis* were considered distinct species, based on the presence of spines on resting spores of *C. radians* and their absence in spores of *C. socialis*. However, Proschkina-Lavrenko (1963) combined both in the same species, *C. socialis* becoming *C. socialis* f. *socialis* and *C. radians* becoming *C. socialis* var. *radians*.

Molecular techniques have led to the possibility of species delimitation based on genetic markers. One of such studies carried out by Kooistra *et al.* (2010) provided the first insights into the phylogeny of the family Chaetocerotaceae, based on the nuclear encoded large ribosomal subunit (LSU or partial 28S rRNA). This work revealed the presence of cryptic diversity in the genus *Chaetoceros*. The presence of cryptic diversity was shown for *C. socialis*, *C. peruvianus*, *C. debilis* and *C. curvisetus*. In the case of *C. socialis*, the strains

from the Gulf of Naples (Italy) were genetically distinct from strain CCMP172 from Friday Harbour, San Juan Islands (North Pacific). However, the morphological description of this latter strain was not documented.

Later, Degerlund *et al.* (2012) studied *C. socialis* strains isolated from the Atlantic/Arctic (northern strains) and the Tyrrhenian Sea (southern strains), examining their morphology, physiology (growth rate, photosynthetic yield) and molecular information (partial 28S rRNA). Clear differences were observed between Atlantic/Arctic and Tyrrhenian strains in spore morphology and physiological traits; strains of different geographic origins were also genetically distinct. The authors did not mention any distinct morphological differences in the vegetative cells. However, the spores of the Atlantic/Arctic strains were reported as smooth, while the strains from the Tyrrhenian Sea produced spiny spores. Based on the above information, it was suggested that the strains from different geographic regions were different species, one inhabiting cold waters and the other warmer waters, but the morphological characterisation was not complete. This interpretation was further supported by Huseby *et al.* (2012), based on distinct metabolomics fingerprinting.

A detailed morphological and molecular analysis of *C. socialis* complex was carried out by Chamnansinp *et al.* (2013). The authors amended the description of *C. socialis,* based on epitype material collected from warm waters of Japan, whereas the species collected from colder (*C. radians*) waters was described as *C. gelidus* Chamnansinp, Li, Lundholm & Moestrup. A few morphological characters were found that differentiate the two species. *Chaetoceros gelidus* has narrower apertures between the cells and the setae emerge very close to the margin of the cells while in *C. socialis sensu stricto* they emerge more inside the valve. The long seta of *C. gelidus* is smooth, with short spines at its distal part where it joins with other long setae of adjacent cells, whereas in *C. socialis* the long seta is covered over its whole length by small spirally arranged spines and poroids. Finally, the resting spores of

C. socialis are spiny compared to smooth resting spores in *C. gelidus*. Chamnansinp *et al.* (2013) also analysed strain CCMP172 (used by Kooistra *et al.* 2010), which, based on morphological and molecular information, was considered to occupy an intermediate position between *C. gelidus* and *C. socialis*. However, since the morphology of the resting spore was not documented for CCMP172, the authors considered that it was premature to describe it as a separate species, although they suggested that the CCMP172 strain could be a third species of the *C. socialis* complex.

In the present chapter, I provide a phenotypic and genotypic characterisation of two new species in the *C. socialis* complex, *C. sporotruncatus* sp. nov., and *C. dichatoensis*.sp. nov., which were collected from the Chilean coast (Las Cruces and off Concepción). From the phylogenetic analysis, it was revealed that the CCMP172 strain was identical to the *C. sporotruncatus*. I also include morphological and genetic information on *C. radicans* and *C. cinctus* strains, also isolated from the Chilean coasts, that turned out to be sister clades to the *C. socialis* complex and have somewhat similar morphology.

Materials and Methods

Phylogenetic analysis: all sequences assigned as *C. socialis* and *C. gelidus* in GenBank were retrieved and included in the alignment. The final alignment comprised the sequences of *C. socialis sensu stricto*, *C. gelidus*, *C. sporotruncatus* and *C. dichatoensis*, along with the sequences of *C. radicans* and *C. cinctus*. Sequences of *C. costatus*, *C. tenuissimus, C. neogracile* and *C. calcitrans* were used as outgroups based on previous phylogenetic information (Kooistra *et al.* 2010; Degerlund *et al.* 2012; Chamnansinp *et al.* 2013). For *C. socialis sensu stricto*, only a few representative sequences from different geographic locations were considered for the analysis, as most of the sequences were identical and would not add phylogenetic information.

Table 5.1. Strains and sequences of Chaetoceros species included in this study. Species name, strain code, collection site, collection date, and morphological data (LM, SEM, TEM); GenBank accession number of 18S and 28S rDNA sequences. An asterisk marks the type strain of the different species. NA = Not available. "#" indicate sequences used for phylogenetic analysis.

Species name	Strain code	Collection site	Collection date	Morphology images	GenBank 18S	GenBank 28S
C. socialis	YL1*	China	15 Mar. 2009	Chamnansinp et al. 2013	NA	# KF219701
	No_1	Thailand	31 Oct. 2008	Chamnansinp et al. 2013	NA	# KF219700
	KMMCC B734	South Korea	6 Jul. 2004	NA	# JQ217339	# JQ217339
	RR	NA	NA	NA	# AY485446	NA
	newLC4	Gulf of Naples, Italy	16 Apr. 2013	NA	# KY852274	As KY852295
	newEA3	Gulf of Naples, Italy	16 Apr. 2013	NA	# KY852277	As KY852295
	newCB4	Gulf of Naples, Italy	16 Apr. 2013	NA	# KY852276	# KY852295
	newEC4	Gulf of Naples, Italy	16 Apr. 2013	NA	# KY852275	# KY852296
	Na2C4	Gulf of Naples, Italy	26 Nov. 2013	LM, SEM	NA	# KY852293
	Na4C4	Gulf of Naples, Italy	26 Nov. 2013	LM, SEM	NA	As KY852295
	Na12C2	Gulf of Naples, Italy	19 Mar. 2014	LM	NA	As KY852295
	Na33B1	Gulf of Naples, Italy	14 Jul. 2015	LM	NA	As KY852295
	Na33C3	Gulf of Naples, Italy	14 Jul. 2015	LM	NA	As KY852295
	Ro4A1	Roscoff, France	11 Aug. 2014	LM	NA	As KY852294
	Ro4A2	Roscoff, France	11 Aug. 2014	LM	NA	# KY852294
C. gelidus	CNCIII51_13	Central Arctic	NA	NA	# HM581777	NA
	WC-13-MC1138	Bering Sea, Pacific Ocean	23 Mar. 2010	NA	# KC771205	NA
	RCC1992	Beaufort Sea	Jun. 2009	NA	# JF794042	NA
	D8*	Skovshoved Harbour, Denmark	Jun. 2009	Chamnansinp et al. 2013	NA	# KF219703
	CPH22	Denmark	16 Dec. 2010	Chamnansinp et al. 2013	NA	# KF219714
	AMB-66	Finnmark coast, Norway	NA	NA	NA	# HE573573
	RCC1990	Beaufort Sea	Jul. 2009	NA	NA	# JQ995407
	E65PG4	Beaufort Sea	Jun. 2009	NA	NA	# JQ995393

Species name	Strain code	Collection site	Collection date	Morphology images	GenBank 18S	GenBank 28S
C. sporotruncatus	Ch2A4*	Las Cruces, Chile	16 Oct. 2013	LM, SEM	# KY852270	# KY852297
	Ch9C4	Concepción, Chile	29 Oct. 2013	LM, SEM, TEM	NA	# KY852298
	CCMP172	San Juan Island, USA	NA	NA	NA	# EF423466
	UNBF-P38C1	New Brunswick, Canada	22 Jun. 2010	NA	NA	# KC986102
C. dichatoensis	Ch4A4*	Las Cruces, Chile	16 Oct. 2013	LM, SEM, TEM	# KY852271	# KY852300
	Ch1B3	Las Cruces, Chile	16 Oct. 2013	LM, SEM	# KY852272	# KY852299
	Ch9B4	Concepción, Chile	29 Oct. 2013	LM, SEM	# KY852273	# KY852301
C. radicans	Ch1B4	Las Cruces, Chile	16 Oct. 2013	LM, SEM, TEM	# KY852260	# KY852288
	Ch2A2	Las Cruces Chile	16 Oct. 2013	LM	# KY852259	# KY852289
	Ch3B1	Las Cruces, Chile	16 Oct. 2013	LM	# KY852261	# KY852290
	Ch10A3	Las Cruces, Chile	29 Oct. 2013	LM, SEM	# KY852263	# KY852291
	Ch11A4	Las Cruces, Chile	1 Nov. 2013	LM	# KY852262	# KY852292
	CCMP197	Fladen Ground, North Sea	1 Jan. 1977	NA	# AB430592	# AB430626
C. cinctus	Ch3A1	Las Cruces, Chile	16 Oct. 2013	LM	# KY852265	# KY852281
	Ch3C4	Las Cruces, Chile	16 Oct. 2013	LM	# KY852266	# KY852282
	Ch6A2	Las Cruces, Chile	16 Oct. 2013	LM	# KY852264	# KY852283
	Ch10B1	Las Cruces, Chile	29 Oct. 2013	LM	# KY852267	# KY852284
	Ch10B3	Las Cruces, Chile	29 Oct. 2013	LM, SEM, TEM	# KY852268	# KY852285
	Ch10B4	Las Cruces, Chile	29 Oct. 2013	LM	NA	# KY852286
	Ch11C3	Las Cruces, Chile	1 Nov. 2013	LM	# KY852269	# KY852287
C. costatus	Na32B1	Gulf of Naples, Italy	22 May. 2015	Kooistra et al. 2010	# KY852258	# KY852280
C. tenuissimus	newCA3	Gulf of Naples, Italy	16 Apr. 2013	Kooistra et al. 2010	# KY852257	# KY852279
C. neogracilis	RCC2275	Beaufort Sea	Jul. 2009	NA	# JN934684	# JQ995449
C. calcitrans	CCMP1315	NA	1960	NA	# KY852256	# KY852278

Distance based neighbor-joining (NJ) method and character based maximum likelihood (ML) maximum parsimony (MP) methods and Bayesian inference (BI) were used to infer the phylogenies. NJ and MP analysis were performed using MEGA v6.06 (Tamura *et al.* 2013). The NJ tree was set to produce a bootstrap consensus tree using Kimura-2-parameter model with 1000 replicates. MP analysis was done using heuristic searches with random addition of sequences (1000 replicates) and a branch swapping algorithm i.e. tree-bisection-reconnection. Gaps were treated as missing data and characters treated as multistate and unordered. ML analysis was performed using RAxML as implemented in raxmlGUI v.1.5beta (Silvestro & Michalak 2012). ML trees were generated using a GTRGAMMA model of base substitution with 'through bootstrap analysis'. Bootstrap values associated with internodes were based on 1000 replicates, with each replicate assigned 10 runs. The ML analysis was performed with the branch-swapping algorithm, through tree-bisection-reconnection. Bayesian trees were constructed using MrBayes 3.2.2 on XSEDE (Ronquist & Huelsenbeck 2003) with GTR+r+PINVAR parameters being estimated during the run and using the default value of four Markov chains. The 'temperature' parameter was set to 0.2, resulting in incremental heating of each chain. The Monte Carlo Markov Chain (MCMC) length was set to 1 million generations, with a posterior probability of bipartitions (i.e. split frequencies) sampled every 100 generations and diagnosed every 1000 generations. For my analysis, I remove the initial 25% of the sampled trees as burn-in. The BI-consensus trees were generated from the sampled trees.

DNA polymorphism: nucleotide composition and substitution patterns were calculated using the MEGA v6.06 (Tamura *et al.* 2013) programs. Distance estimations for nucleotides were obtained using the Kimura-2-Parameter model, with the rate of transitional/transversional substitution. The standard deviation (SD) was estimated using the bootstrap method with 1000 replicates.

Results

Phylogenetic analysis: an overview of the phylogenetic relationships is illustrated in Fig. 5.2. for 18S and Fig. 5.3. for 28S. All the phylogenetic analyses resulted in a similar clade topology for ML, BI, and MP for 18S and 28S; hence, I used the ML tree to describe the clades. With sequences of *C. tenuissimus, C.* cf. *neogracile* and *C. calcitrans* designated as an outgroup, the sequences of *C. costatus* resolved as the first clade to branch off with high bootstrap values in both the 18S and 28S trees.

18S phylogeny: a total of 215 MP trees were retained for length 523, a rescaled consistency index 0.730, a homoplasy index 0.210, and a G-fit of –223.650. For the inference of the ML tree, the chosen model was GTR with a base substitution rate matrix of A↔C (0.9967), A↔G (2.7786), A↔T (1.0760), C↔G (0.6425), C↔T (4.7611), relative to G↔T (1.0000); gamma-distributed rate variation among sites with shape parameter α=0.1332 and estimated base frequencies A=0.273, C=0.190, G=0.251, and T=0.286. The likelihood score (-ln L) for the ML tree for 18S alignment was 5320.9492. Bayesian inference was conducted with one million MCMC runs. The posterior probability of the sampled trees showed a standard deviation value of 0.003. After discarding 25% of the initially sampled trees as burn-in, and 7501 trees were sampled for my inference.

The inferred ML phylogeny (Fig. 5.2) revealed two distinct sister clades with high bootstrap support, one clade comprising the *C. socialis complex* (90%) and the other with *C. radicans* and *C. cinctus* (100%). Sequences of *C. radicans* and *C. cinctus* resolved as sister clades to the clade including the *C. socialis complex*. *Chaetoceros radicans* and *C. cinctus* formed two clades supported with high bootstrap values (100%). Within the *C. radicans* clade, two clades were resolved; one comprised *C. radicans* sequences (100%) from Chile and the other one represented by a CCMP197 sequence from the Fladen Ground, North Sea (100%).

Although the 28S rDNA sequence of CCMP197 was incomplete (only 586 bp), it still showed four nucleotide differences over the region covered.

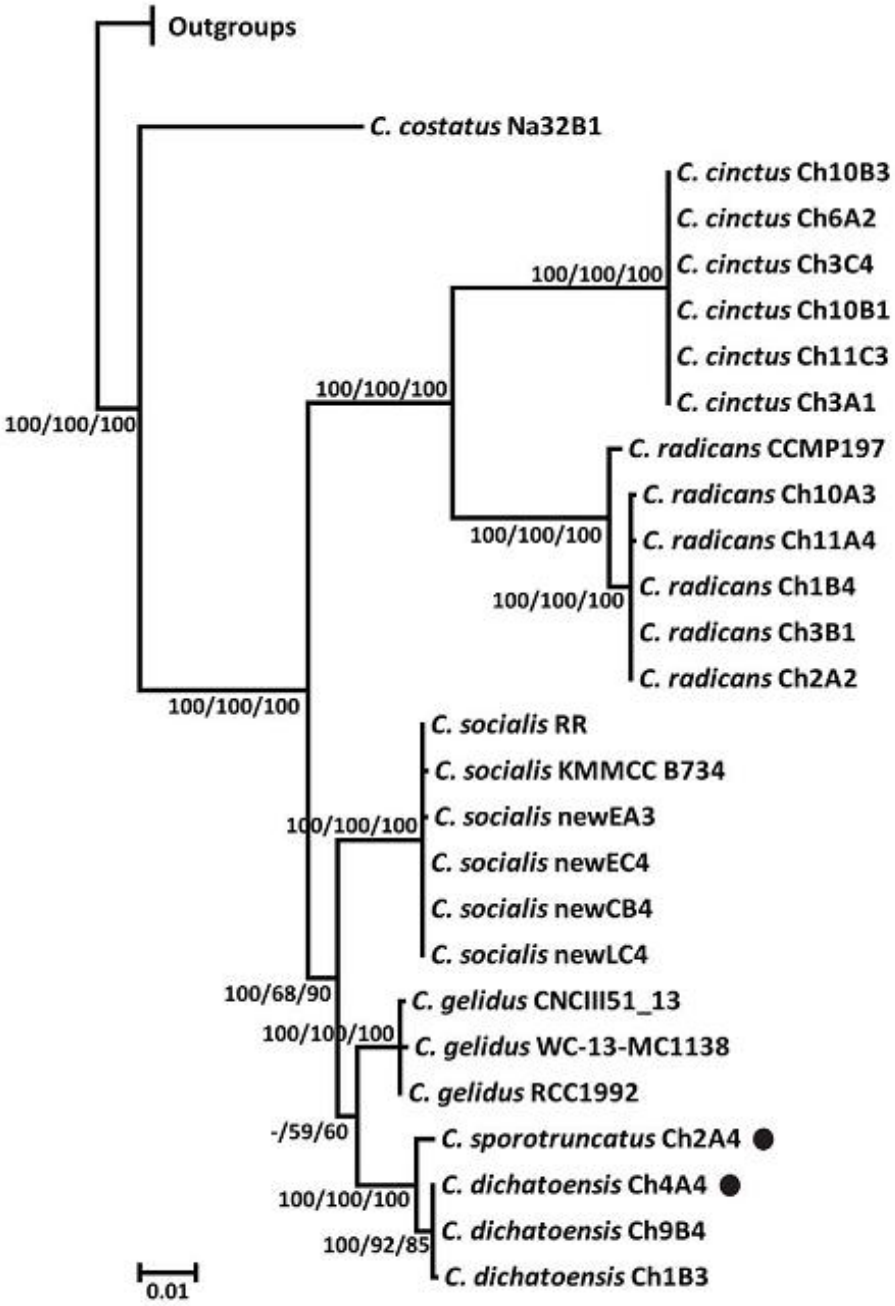


Fig. 5.2. A maximum likelihood phylogeny inferred from the 18S rDNA sequences of strains belonging to the species of *Chaetoceros socialis* complex and its close relatives, *C. radicans* and *C. cinctus*. Far outgroups have been pruned away from the tree topology. Numbers on nodes denote bootstrap supports and Bayesian posterior probabilities for maximum likelihood, maximum parsimony, and Bayesian inference, respectively. Black dots indicate the type strains of the species in the *C. socialis* complex (18S sequences are not available for the type strains of *C. socialis* and *C. gelidus*.

In the clade containing the *C. socialis* complex, four terminal clades were detected. The first clade included sequences of *C. socialis sensu stricto* (100%) including the sequence of strain designated as type material (strain YL1 from South China sea; Chamnansinp *et al.* 2013). Sequences of strains isolated in this study from the Gulf of Naples (Italy) and Roscoff (France) clustered in this clade as well. Some GenBank sequence from Thailand, China, and Denmark also grouped in this clade. The next, well-supported clade was the one including a clade (100%) with *C. gelidus* as sister to a clade with the two new species, though this clade obtained only weak support (60%). The clades of two newly described species, *C. sporotruncatus,* and *C. dichatoensis* obtained good support (100% viz 85%, respectively).

The resulting BI tree also revealed a similar tree topology which corroborates with that of the ML with a high Bayesian posterior probability (bpp). In comparison, the MP tree also showed similar topology, but the clade with *C. socialis* complex was weakly supported along with the clade (59%) containing *C. gelidus* and the two new species. The MP and BI trees were overlapped on the ML tree and the valves are depicted in percentage.

28S phylogeny: a total of 300 MP trees were retained for length 496, a rescaled consistency index 0.689, a homoplasy index 0.252, and a G-fit of –164.600. For the inference of the ML tree, the chosen model was GTR with a base substitution rate matrix of A↔C (0.7079), A↔G (2.1148), A↔T (0.8197), C↔G (0.7833), C↔T (3.8743), relative to G↔T (1.0000); gamma-distributed rate variation among sites with shape parameter α=0.3670 and estimated base frequencies A=0.279, C=0.186, G=0.276, and T=0.259. The likelihood score (-ln L) for the ML tree for 18S alignment was 3383.7855. Bayesian inference was conducted with one million MCMC runs. The posterior probability of the sampled trees showed a standard deviation value of 0.006. After discarding 25% of the initially sampled trees as burn-in, and 7501 trees were sampled for my inference.

The inferred ML tree showed a topology similar to that in the 18S tree (Fig. 5.3). In contrast to the 18S tree, sister relationships between the clade including all sequences of *C. radicans* (100%) and the one with *C. cinctus* (100%) did not obtain support. Instead, the clade of *C. cinctus* branched off first and only then the clade with *C. radicans*, though this relationship did not obtain any bootstrap support. As in the 18S tree, the sequence of *C. radicans* CCMP197 was distinct from the sequences obtained from Chilean strains assigned to this species. In contrast to the situation in the 18S tree, the clade with sequences of the *C. socialis* complex (*C. socialis* from Naples, *C. gelidus* and *C. sporotruncatus* and *C. dichatoensis* from Chile) was weakly supported (74%).

Within the clade of the *C. socialis* complex, four well-supported clades were resolved (≥96%). As in the 18S tree, the first to branch off was the clade with *C. socialis sensu stricto*. Its sister clade, which included the clade with *C. gelidus,* and the clade containing the two clades of the new species (*C. sporotruncatus* sp. nov. and *C. dichatoensis* sp. nov.) obtained very weak support (53%). Within the *C. socialis sensu stricto* clade, I found that the sequence from Thailand (KF219700) was identical except for the five base changes (C↔G, G↔C, T↔C, C↔T, and C↔G); and the sequence obtained from the South China Sea (KF219701) differed by one base change (G↔A).

The inferred BI tree revealed two distinct sister clades; one clade comprising the *C. socialis complex* (99%) and the other with *C. radicans* and *C. cinctus* (63%). Sequences of *C. radicans* and *C. cinctus* resolved as sister clades (as seen in 18S) to the clade including the *C. socialis complex*. Within the *C. radicans* clade, two clades were resolved; one comprised *C. radicans* sequences (100%) from Chile and the other one represented by a CCMP197 sequence from the Fladen Ground, North Sea (100%). The clades were well supported and revealed a similar clade structure as in ML tree. In contrast, the MP tree showed a basal

polytomy for the two sister clades. The internal topology within *C. socialis* complex was not well supported. The terminal clades were well supported in all the three phylogenetic trees.

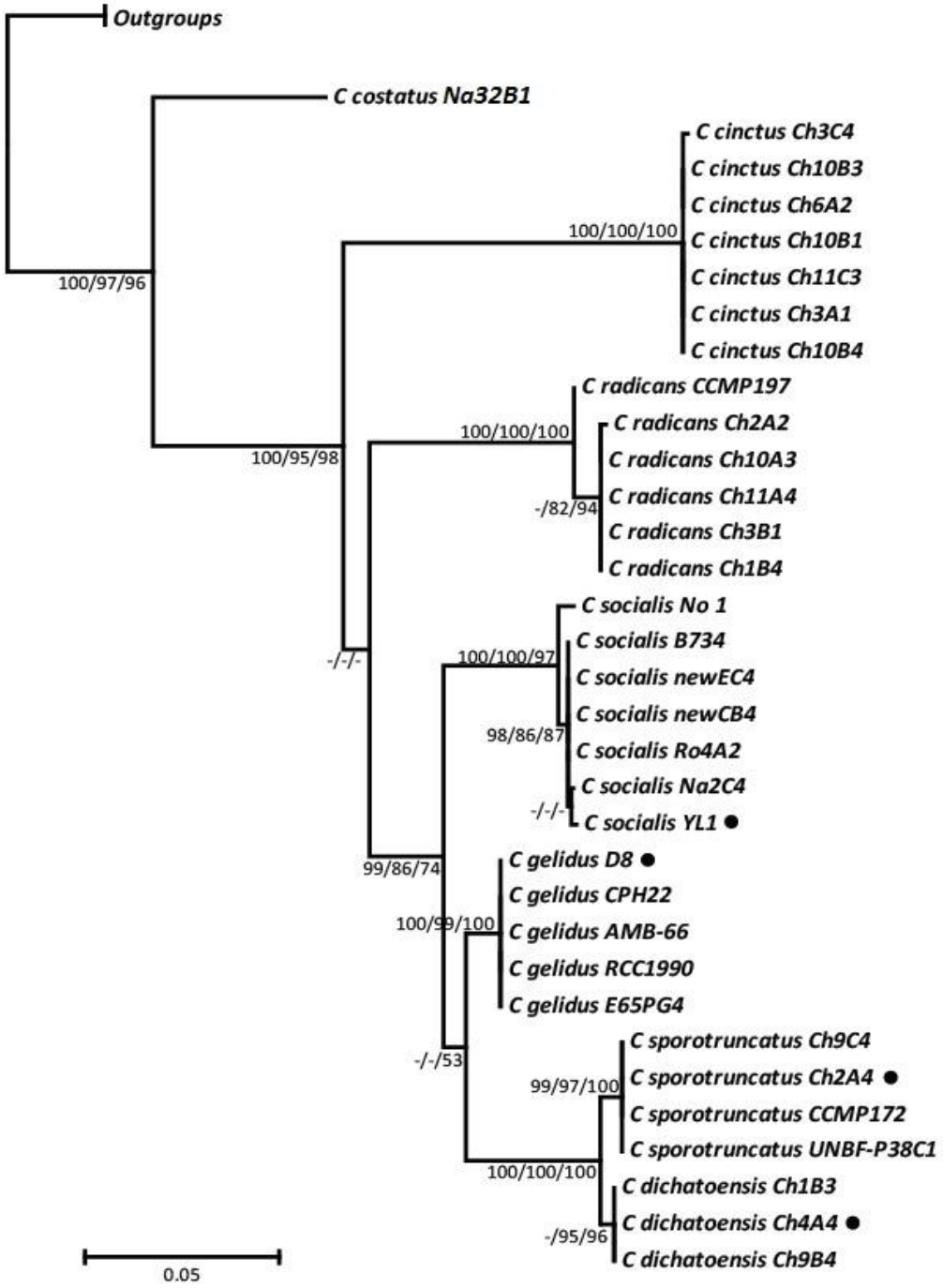


Fig. 5.3. Maximum likelihood phylogeny inferred from the partial 28S rDNA sequences of strains belonging to species of the *Chaetoceros socialis* complex was well as its close relatives, *C. radicans* and *C. cinctus*. Far outgroups have been pruned away from the tree topology. Numbers on nodes denote bootstrap supports and Bayesian posterior probability for maximum likelihood, maximum parsimony and Bayesian inference, respectively. Black dots indicate the type strains of the species in the *C. socialis* complex.

Intraspecific diversity: only the 28S rDNA alignment was used to estimate the nucleotide distance among the species used in the phylogeny. The pairwise nucleotide distance between the different strains used in the phylogenetic analyses is presented in Table 5.2. Only those sequences that showed few variations among the species were used to estimate the nucleotide distance using Kimura-2-parameter.

Table 5.2. Pairwise distance matrix based on nucleotide sequences of *Chaetoceros socialis* complex. Pairwise distances calculations were accomplished using MEGA v6.06 (Tamura et al. 2013) software according to Kimura-2-P model, and Standard Deviation (SD) estimated by the bootstrap method (above the diagonal). Distances and standard errors are displayed in the lower-left matrix and the upper-right matrix, respectively.

Pairwise Distance (K2P)\SD	*Chaetoceros socialis* Na2C4	*Chaetoceros socialis* YL1	*Chaetoceros socialis* KF219700	*Chaetoceros gelidus* RCC1990	*Chaetoceros sporotruncatus* Ch2A4	*Chaetoceros dichatoensis* Ch4A4
***C. socialis* Na2C4**		0.001	0.003	0.008	0.010	0.010
***C. socialis* YL1**	0.001		0.003	0.008	0.010	0.010
***C. socialis* KF219700**	0.007	0.008		0.008	0.010	0.010
***C. gelidus* RCC1990**	0.047	0.049	0.049		0.008	0.007
***C. sporotruncatus* Ch2A4**	0.070	0.072	0.072	0.047		0.004
***C. dichatoensis* Ch4A4**	0.069	0.070	0.070	0.044	0.009	

The genetic data confirmed the distinction of the *C. socialis sensu stricto* and *C. gelidus* from the other two Chilean species. The overall average estimated nucleotide distance (according to K2P model) between the *C. socialis* complex was 0.048 (on an average length of 760 bp alignment; see Table 5.2). The estimated pairwise distance between the two new species from Chile was 0.009 (corresponding to seven nucleotide position, in the 760 bp alignment, refer Table 5.2). The sequence of *C. sporotruncatus* from Chile and CCMP172 from San Juan Island, USA did not show differences between these species. Whereas, within the *C. socialis sensu stricto* sequences collected from Thailand showed a nucleotide distance of 0.007 (corresponding to the five nucleotide position,) and South China Sea sequence with 0.001 (corresponding to the five nucleotide position) in the 760 bp of the 28S alignment.

Morphology of the species included in the *Chaetoceros socialis* complex

Chaetoceros socialis sensu stricto Lauder

Fig. 5.4. a-l; Table 5.3

I have examined the morphology of five strains of *C. socialis sensu stricto* isolated from the Gulf of Naples and two strains isolated from Roscoff (Table 5.1). The gross morphology and ultrastructure of these strains match the amended description of *C. socialis sensu stricto* (Chamnansinp *et al.* 2013).

Cells join in chains that are curved in broad girdle view (Figs 5.4 a, b). Single cells also were observed in culture. Cells have a single large chloroplast. In girdle view, cells appear quadrangular. The length: width ratio is variable: cells are mostly wider than longer, but also longer than wider cells were observed. Each cell has three curved setae and one long seta that protrude towards the concave part of the chain. The tips of the long setae converge to a common point and connect together with long setae from other chains, forming a curved colony. The curved colonies form a spherical colony (Fig. 5.4. a). Single cells with all short setae were also observed. Setae emerge from the inside of the valve face with a basal part and cross over on the valve margin, thus forming a wide hexagonal aperture (Fig. 5.4. c). All setae are circular in cross section and possess spirally arranged spines and poroids along with large, elongated, solitary pores (Figs 5.4. d–e). The valve face is broadly elliptical or oval. The valve is ornamented with weak radial, or dichotomously branching, costae radiating from a roundish central annulus and from the two insertions points of the setae (Figs 5.4. g–h). The area within the annulus is hyaline (Figs 5.4. g–h). The terminal valve has a single central rimoportula with a short external tube (Fig. 5.4. f); internally, the rimoportula is a slightly elongated slit. No rimoportula is present on the intercalary valves (Figs 5.4. g–h).

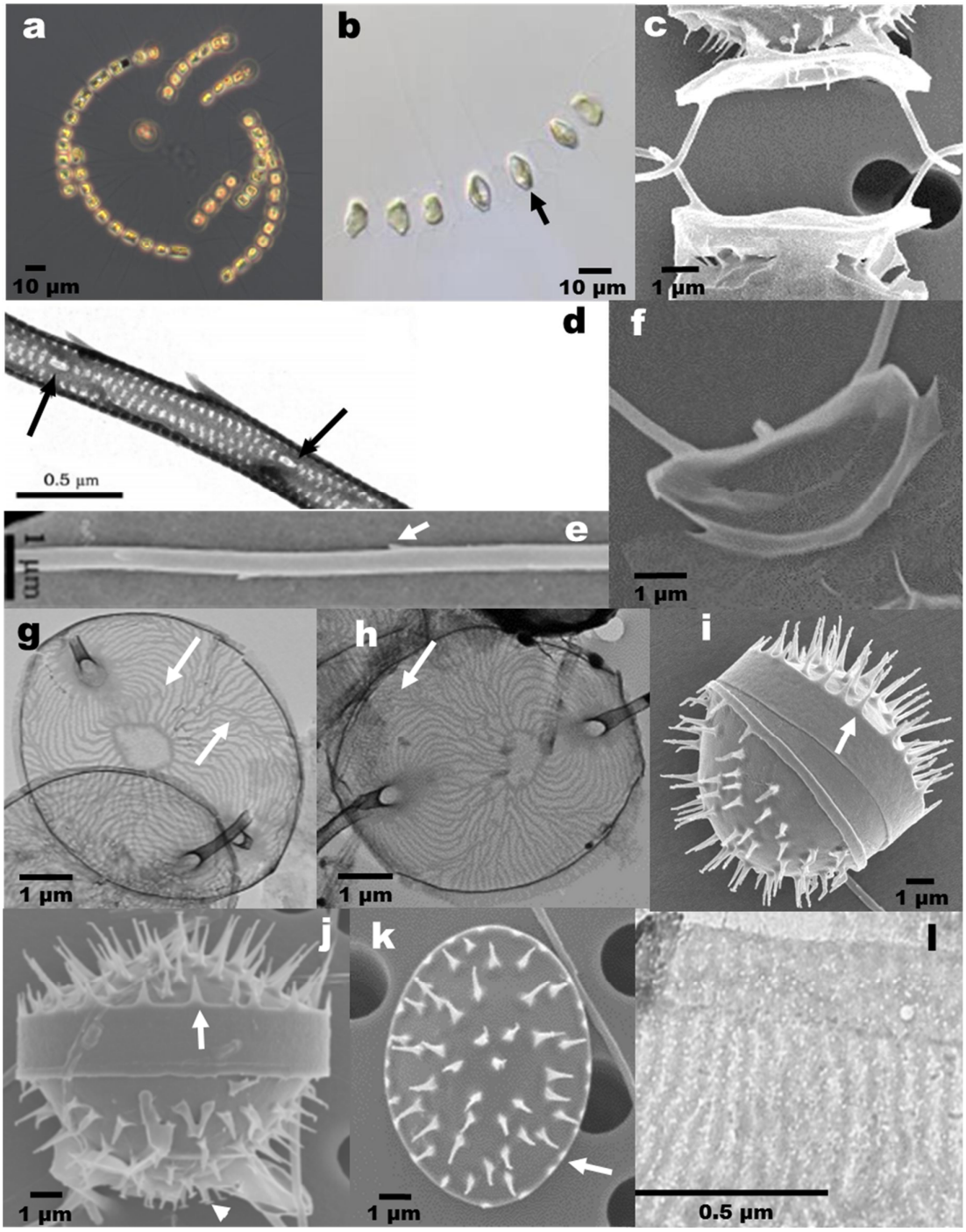


Fig. 5.4. *Chaetoceros socialis sensu stricto*. Strain Na2C4 from Naples, Italy (a, c, g–h and j–l); strain MR27 from Thailand (d–e: from Chamnansinp *et al.* 2013); strain Ro4A2 Roscoff, France (b); strain from Naples, Italy, courtesy of Diana Sarno (f); strain SZN-423 from Naples, Italy from Kooistra *et al.* (2010, i). LM (a-b), SEM (c, e–f and i–k), TEM (d, g–h and i). (a) A spherical colony with multiple chains. (b) Single chain with resting spores in girdle view (arrow). (c) Aperture. (d) A detail of a short seta, with spines and pores; arrows mark large poroids. (e) A detail of a long seta; arrow marks the minute spine. (f) Terminal valve with central rimoportula. (g–h) Intercalary valve with central annulus and branched costae (arrowed). (i–k) Resting spore; a single row of spines on the margin of the primary valve (arrowed in i and j); bifurcated spines at times present (arrowhead in j). (l) Girdle band.

Girdle bands are ornamented with numerous parallel costae and scattered small poroids (Fig. 5.4. l). Apical axis is 11.4±1.3 (8.9−13.3) µm long, pervalvar axis is 8.6±2.0 (6.1−14.4) µm long and size of aperture in pervalvar axis is 4.9±1.5 (2.8−8.9) µm (Table 5.3).

Spores are biconvex; valves are hemispherical-conical, sometimes one valve more convex than the other (Figs 5.4. b and i–j). Spores are covered with stiff spines; the margin of the primary valve has a single row of spines (Figs 5.4. i–k). Spines are conical in shape; at times spines bifurcate at the tip were observed (arrowhead in Fig. 5.4. j).

***Chaetoceros gelidus* Chamnansinp, Li, Lundholm & Moestrup**

Fig. 5.5. a–j; Table 5.3

In the present study, I did not isolate strains belonging to *C. gelidus*. In the following, I present a description of the species based on the literature (Degerlund *et al.* 2012; Chamnansinp *et al.* 2013). The illustrations presented in Figs 5.5. were reproduced from Chamnansinp *et al.* (2013).

Cells with a single large chloroplast joined in chains (Figs 5.5. a−b). The long seta of each cell protrudes towards one side of the colony and joins the long setae of adjacent cells, forming a curved chain (Fig. 5.5. b). Several chains group together forming a more or less spherical colony. Cells usually quadrangular in girdle view; they can be longer than wide or wider than long. The aperture is hexagonal, narrow in pervalvar view, with a slight central constriction (Fig. 5.5. f). Setae arise just inside the valve corner with a short basal part and cross over on the chain margins (Fig. 5.5. f). Setae are circular in cross section. Short setae possess sharp spines arranged in a helicoidal pattern and a spiral row of small poroids (Figs 5.5. c and e). The long setae are smooth, with small spirally organized poroids and spines only towards the end where they join with the adjacent setae. Solitary large pores are present on both types of setae (Figs 5.5. d−e). Valves appear broadly elliptical to circular.

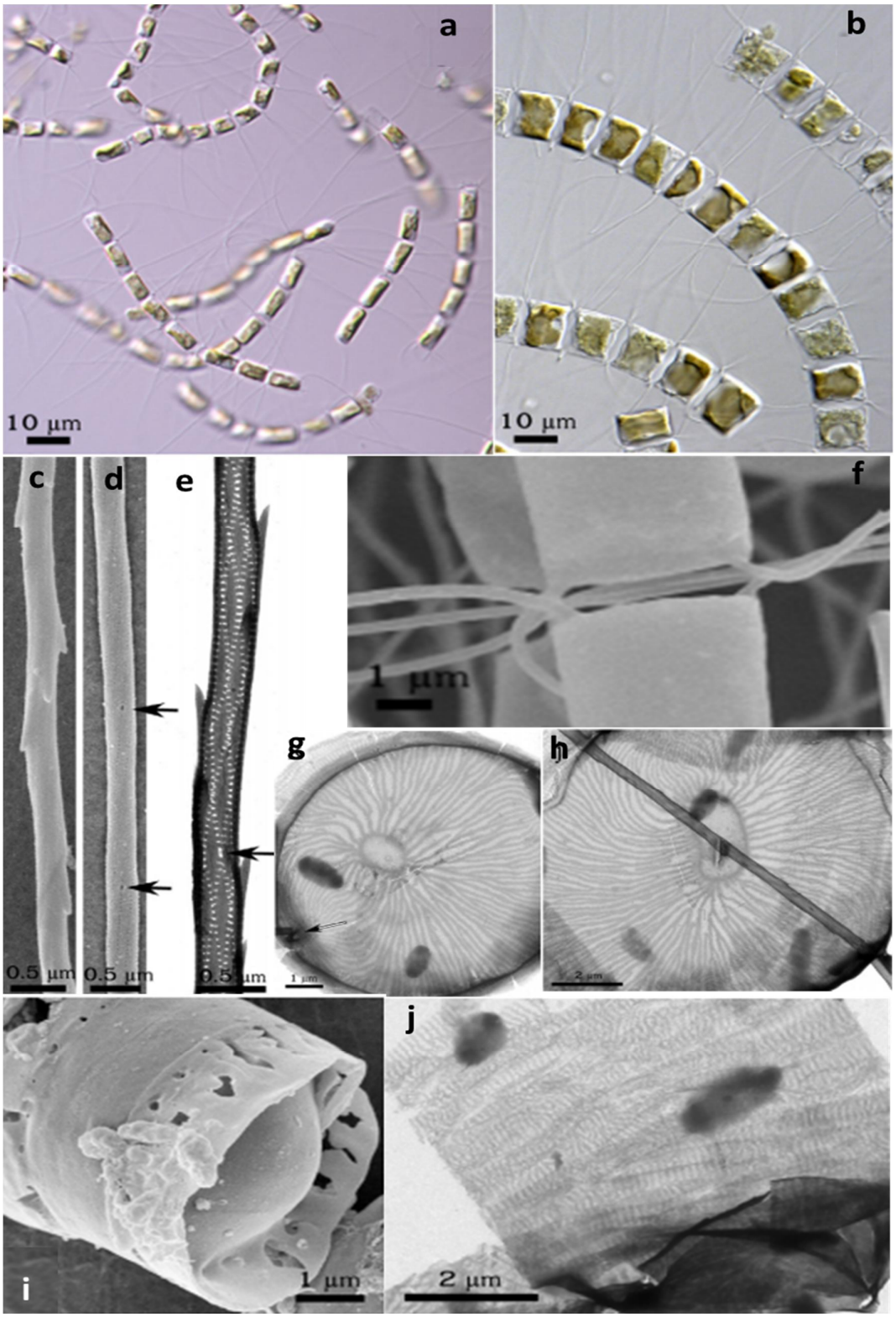


Fig. 5.5. *Chaetoceros gelidus*. The morphological illustrations are taken from Chamnansinp *et al.* (2013). Cultures from Denmark (a, c–i) and Greenland (b). LM (a–b), SEM (c–d, f and i), TEM (e, g–h, and j). (a) Colonies of vegetative cells. (b) Chains in broad girdle view. (c and e) Short setae with spines and rows of poroids (arrow marks a large pore). (d) Long smooth seta with large pores, marked with arrows. (f) Aperture. (g) Intercalary valve with central annulus and branched costae. (h) Terminal valve with central rimoportula. (i) Resting spores with crest. (j) Girdle bands.

Valve ornamented with costae radiating from the central annulus and partly converging towards the two insertions points of the setae (Fig. 5.5. g). The annulus shape is variable, from elliptical to elongate and sometimes irregular. The area within the annulus is hyaline. Terminal valve with a single central rimoportula (Fig. 5.5. h). No rimoportula is present on the intercalary valves. Girdle bands are ornamented with numerous parallel costae and two rows of scattered poroids (Fig. 5.5. j). Apical axis is 7.1±2.7 (3.6−16.9) µm long, the pervalvar axis is 9.7±1.8 (5.0−14.7) µm long, and the size of the aperture in the pervalvar axis is 2.0±0.6 (0.7−3.3) µm. The spores are biconvex, one side usually more convex than the other, and smooth. The mantles of both valves have a crest of more or less fused projections (Fig. 5.5. i). The crest on the hypovalve possesses a series of longitudinal slots.

***Chaetoceros sporotruncatus* sp. nov.**

Fig. 5.6. a–k; Fig. 5.7. a–h; Table 5.3

Species diagnosis: each cell has three curved setae and one long seta that joins with the long setae of the adjacent cells in the chain forming a curved colony. The curved colonies are aggregated in turn to form a spherical colony. Cells quadrangular in girdle view, longer than wide or wider than long. One chloroplast per cell. Apertures between adjacent cells are large, hexagonal, and slightly constricted at the centre. Valves elliptical to circular, with a central hyaline annulus and costae radiating from the annulus and partly converging towards the emerging point of the setae. Terminal valve with a central rimoportula. Setae emerge from the inside of the valve face with a basal part and cross over at the chain margin. The three short setae and the long seta are ornamented with spirally arranged spines and poroids, and with large solitary pores. Spines are absent on the proximal part of the long seta. Girdle bands are ornamented with numerous parallel costae and minute scattered poroids. Apical cell axis is 11.2±2.3 (7.2−15) µm long, pervalvar axis 11.3±4.13 (7.2−27.8) µm long and length of the aperture in pervalvar axis 4.6±1.3 (2.2−6.7) µm. Spores biconvex; valves with

the shape of a truncated cone. The primary valve is ornamented in its apical portion with raised lenticular-shaped structures.

Holotype: the TEM grids of strain Ch2A4, isolated at Las Cruces, Chile, deposited at the Museum of Stazione Zoologica Anton Dohrn, Naples, Italy.

Isotype: the SEM stubs of the same strain Ch2A4 deposited at the Museum of Stazione Zoologica Anton Dohrn, Naples, Italy.

Molecular diagnosis: the species is defined by the combined nucleotide sequences of strain Ch2A4: D1-D2 LSU (GenBank KY852297) and SSU (GenBank KY852297).

Type locality: Las Cruces, Chile (33° 29' 46″ S and 71° 37' 39″ W)

Etymology: the species name *sporotruncatus* refers to the shape of the valves, which look like truncated cones.

Detailed description: curved chains of cells forming more or less spherical colonies (Fig. 5.6. a). Single cells were seldom observed in culture. Cells are quadrangular in girdle view (Fig. 5.6. b), longer than wide or wider than long. Each cell has a single large chloroplast (Fig. 5.6. b). Apertures have a wide hexagonal shape, generally with a small central constriction (Fig. 5.6. f) or sometimes flat. Cells with longer apical axis have narrower (in pervalvar axis) apertures. Setae emerge from inside of the valve margin, have a relatively long basal part and crossover of setae occurs at chain margin (Fig. 5.6. f). Cells have three short, curved setae and one longer and straight (Fig. 5.1. a); the long setae of the cells in a chain converge into a common point toward the centre of the curved colony (Fig. 5.6. a). Occasionally, single cells with four setae short have been observed (Fig. 5.1. b).

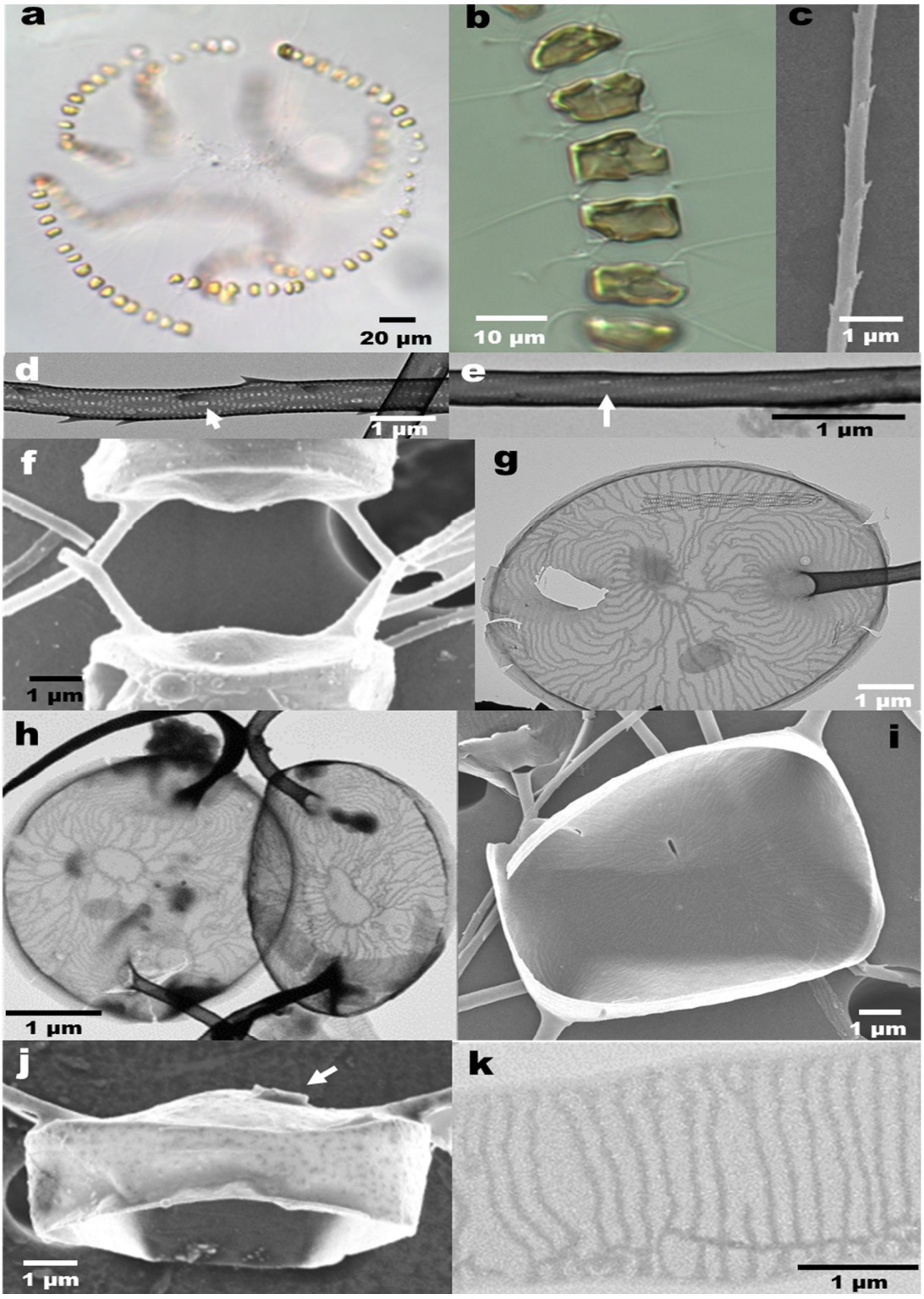


Fig. 5.6. *Chaetoceros sporotruncatus* sp. nov. Strain Ch2A4 from Las Cruces, Chile (a and g–k) and strain Ch9C4 from Concepción, Chile (b–f). LM (a–b), SEM (c, f and i–j), TEM (d-e, g-h and k). (a) A spherical colony with multiple chains. (b) A chain in girdle view. (c) Detail of a short seta with spines. (d) Detail of a short seta with spirally arranged spines and poroids and large pores marked with an arrow. (e) Detail of a long seta; pore marked with an arrow. (f) Aperture. (g–h) Intercalary valves with costae. (i) Internal view of a terminal valve with the slit-shaped central process. (j) External view of the central process (arrow). (k) Girdle band.

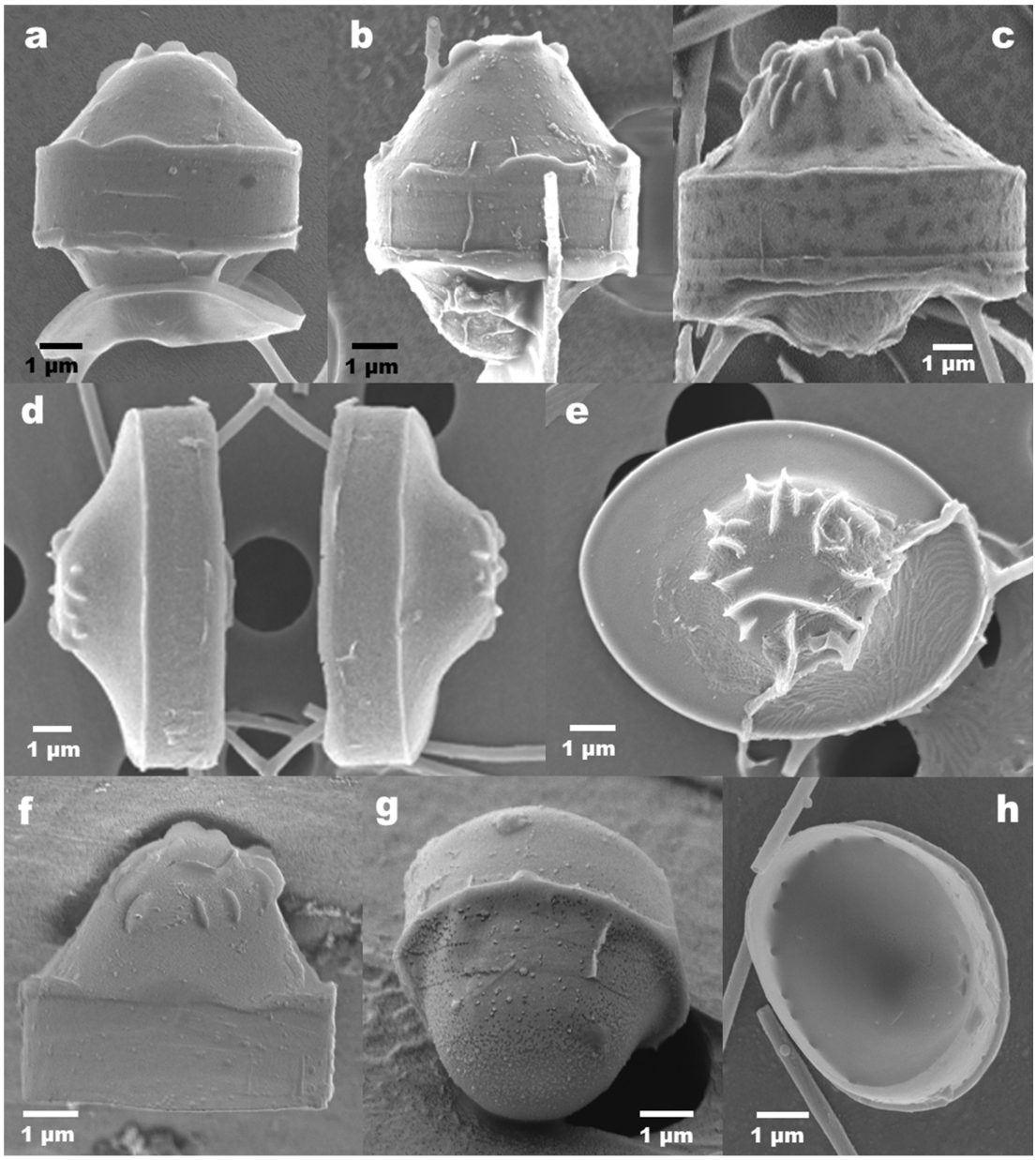


Fig. 5.7. Chaetoceros *sporotruncatus* sp. nov. Strain Ch2A4 from Las Cruces, Chile (c–e) and strain Ch9C4 from Concepción, Chile (a–b and f–h). SEM (a-j). (a–b) Spores, with the primary valve ornamented on its apical part with raised lenticular-shaped structures. (c–d) Partially formed spores; only the primary valve has been formed. (e) Spore in valve view. (f) Primary valve of a spore, lateral view. (g) Secondary valve of a resting spore. (h) Internal view of a secondary valve, note the ring of puncta.

Setae are circular in cross section. Short and long setae have spirally arranged poroids and spines (Figs 5.6. c–e). Large solitary pores are present in both types of setae (Figs 5.6. d–e). Spines are absent on the proximal part of the long seta. Valves are elliptical to circular in outline, ornamented with branching costae radiating from a central annulus and partly converging towards the two insertion points of the setae (Figs 5.6. g–h). The shape of the

annulus is elliptical, at times with an irregular outline; the area within the annulus is hyaline (Figs 5.6. g–h). The terminal valve has a single large central rimoportula. The rimoportula has an elongated slit-shaped opening on the internal side of the valve (Fig. 5.6. i) and a flattened external tube (Fig. 5.6. j). The rimoportula is absent on the intercalary valves. Girdle bands ornamented with numerous parallel costae and scattered minute poroids (Fig. 5.6. k). Apical cell axis is 11.2±2.3 (7.2−15) µm long, pervalvar axis 11.3±4.13 (7.2−27.8) µm long and length of the aperture in pervalvar axis 4.6±1.3 (2.2−6.7) µm. Resting spores are biconvex in shape, with the valves shaped as a truncated cone (Fig. 5.7. a–c). The primary valve is ornamented in its apical portion with raised lenticular-shaped structures of variable size (Figs 5.7. a–f). The secondary valve is smooth (Fig. 5.7. g) and presents a single line of puncta on the margin of the mantle (Fig. 5.7. h).

Chaetoceros dichatoensis sp. nov.

Fig. 5.8 a–j; Fig. 5.9 a–i; Table 5.3

Species diagnosis: each cell has three curved setae and one long seta that joins with the long setae of the adjacent cells in the chain forming a curved colony. The curved colonies form a spherical colony. Cells quadrangular, longer than wide or wider than long. One chloroplast per cell. Apertures between adjacent cells are large, hexagonal in shape and slightly constricted at the centre. Valves elliptical to circular, with a central hyaline annulus and costae radiating from the annulus and partly converging towards the two setae insertions points. Terminal valves with a central rimoportula. Setae emerge from the inside of the valve face with a basal part and cross over at chain margin. The three short setae and the long seta are ornamented with spirally arranged spines and poroids, and with large solitary pores. Spines are absent on the proximal part of the long seta. Girdle bands ornamented with numerous parallel costae and scattered minute poroids. Apical axis is 6.4±1.4 (3.9−10.0) µm long, pervalvar axis is11.6±2.7 (6.1−22.2) µm long and size of the aperture in the pervalvar

axis is 2.6±0.9 (1.1−-6.7) µm. Spores biconvex, with roughly hemispherical valves. Valves ornamented with spines of variable density and size; the basal portion of the spines is at times expanded, forming more or less extended ridges.

Holotype: the TEM grids of the strain Ch4A4 isolated at Las Cruces, Chile, deposited at the Museum of Stazione Zoologica Anton Dohrn, Naples, Italy.

Isotype: the SEM stubs of the same strain Ch4A4 isolated at Las Cruces, Chile, deposited at the Museum of Stazione Zoologica Anton Dohrn, Naples, Italy.

Molecular diagnosis: the species is defined by the combined nucleotide sequences of strain Ch4A4 D1-D2 LSU (GenBank KY852300) and SSU (GenBank KY852271).

Type locality: Las Cruces, Chile (33° 29' 46″ S and 71° 37' 39″ W)

Etymology: the species name *dichatoensis* is dedicated to the Chilean Marine Stations in Di Chato that was destroyed by the tsunami in Feb. 2010.

Detailed description: chains are curved in broad girdle view and aggregate to form a secondary colony, which is spherical (Figs 5.8. a–b). Single cells were at times observed in culture. Cells are quadrangular in girdle view (Fig. 5.8. b). Cells are longer than wide or wider than long. Each cell has a single large chloroplast. Apertures are wide in pervalvar axis and hexagonal (seldom narrower in enlarged cells), generally with a central constriction (Fig. 5.8. d) or sometimes flat. Setae are circular in cross section. Cells have three short and curved setae and one long straight seta; solitary cells with all short setae have been observed. Setae emerge from the inside of the valve margin with a basal part and fusion of setae occurs on the cell margin. All setae possess a spiral row of small poroids together with spirally arranged spines (Figs 5.8. e–g); large solitary scattered pores are also present. Spines are absent on the proximal part of the long seta. The valve face is elliptical or circular. The valves are ornamented with costae radiating from the central annulus and partly converging

towards the two insertions points of the setae (Fig. 5.8. h). The annulus is elliptical or sometimes irregular in shape and the area within the annulus is hyaline (Fig. 5.8. h). The terminal valve has a central rimoportula. The rimoportula has a short external tube (Fig. 5.8. i) and a slit-shape on the inside of the valve (Fig. 5.8. j). The rimoportula is absent on the intercalary valves. The girdle bands are ornamented by numerous parallel costae and scattered minute poroids (Fig. 5.9. i). Apical axis is 6.4±1.4 (3.9−10.0) µm long, pervalvar axis is11.6±2.7 (6.1−22.2) µm long and size of the aperture in the pervalvar axis is 2.6±0.9 (1.1−6.7) µm (Table 5.3). Spores are biconvex in shape, with more or less hemispherical valves (Figs 5.9. a−f). Both the primary and the secondary valves are covered with spines of variable length and density (Figs 5.9. a−g). The basal portion of the spines can be considerably expanded, forming more or less extended ridges that connect different spines (Fig. 5.9. h). A single row of puncta is present on the mantle of the secondary valve (Fig. 5.9. g).

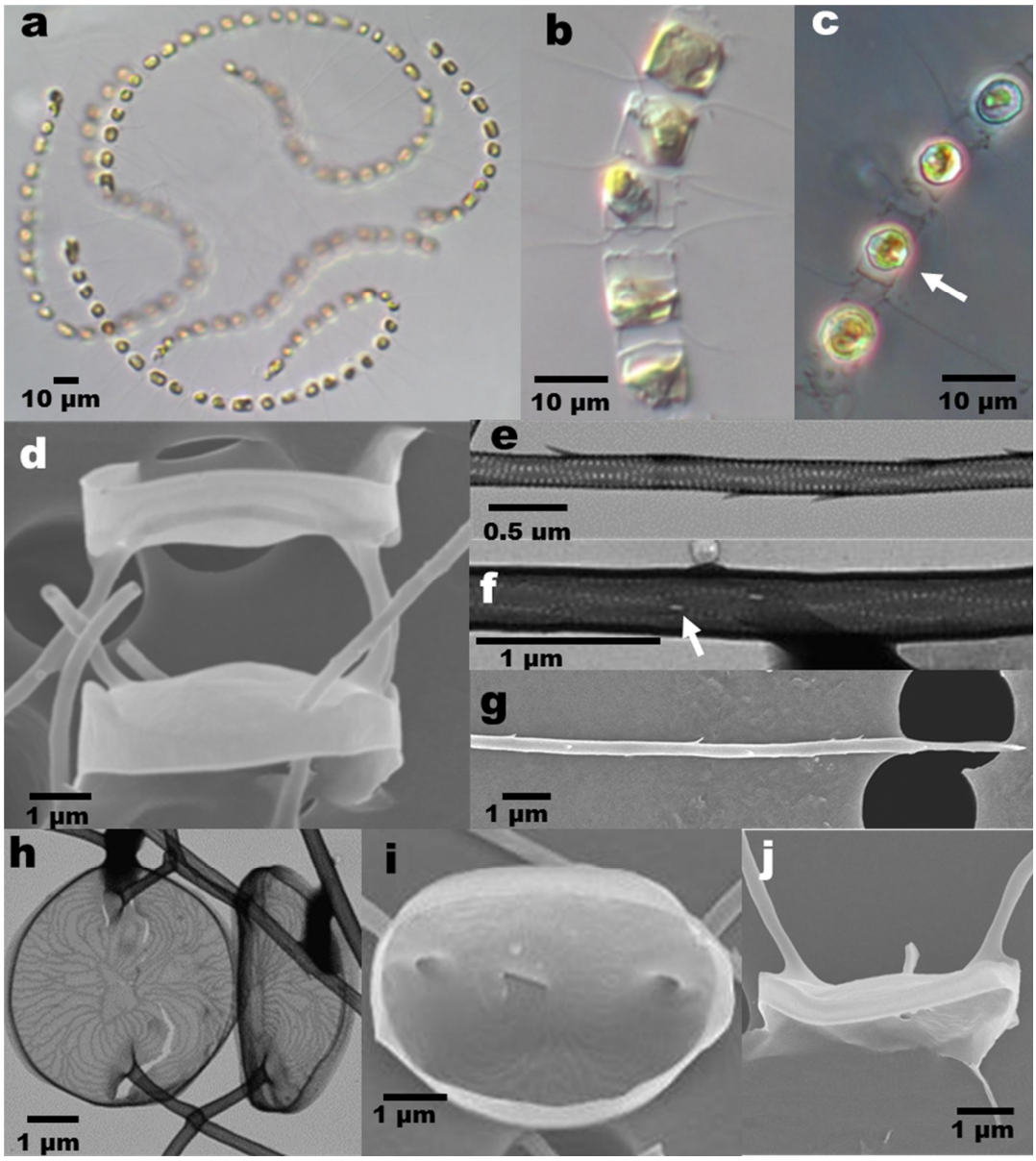


Fig. 5.8. *Chaetoceros dichatoensis* sp. nov. Strain Ch4A4 from Las Cruces, Chile, (a−i). LM (a−c), SEM (d and g−j), TEM (e−f and h). (a) A spherical colony with multiple chains. (b) A chain of cells in broad girdle view. (c) Spores (arrow) in a chain. (d) Aperture. (e) Detail of a seta with spirally arranged rows of spines and poroids. (f) Middle part of a long seta with spirally arranged poroids and pores (arrowed). (g) Terminal part of a long seta with spirally arranged spines. (h) Intercalary valve with costae and annulus. (i) Internal view of a terminal valve with central slit-shaped rimoportula. (j) External view of a terminal valve with a tube-like rimoportula.

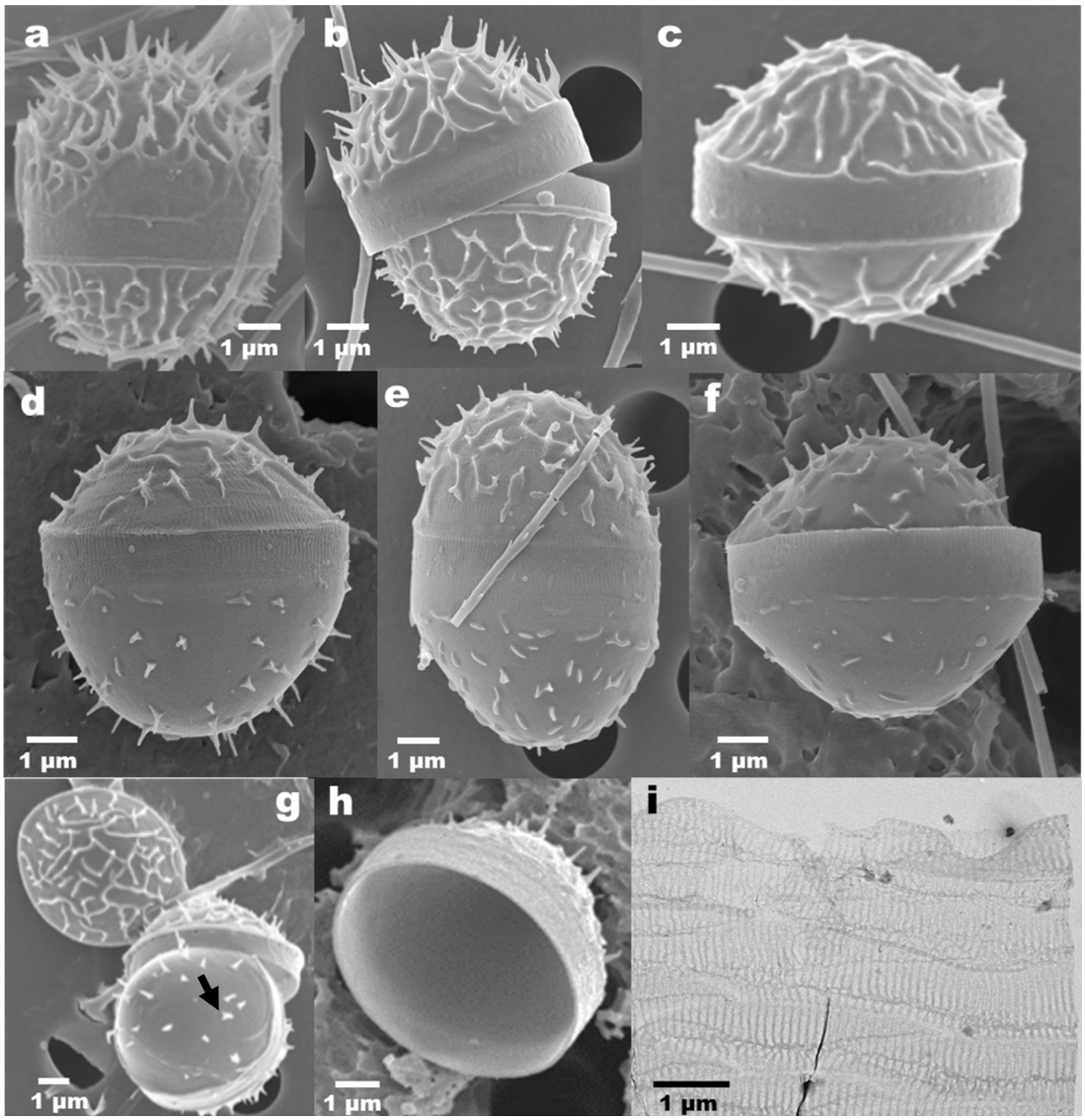


Fig. 5.9. *Chaetoceros dichatoensis* sp. nov. Strain Ch4A4 from Las Cruces, Chile (a−n). SEM (a−h) and TEM (i). (a−h) Resting spores; note the variability of spine size, shape, and density. (g) A line of pores (arrowed) on the mantle of the secondary valve of the spore. (h) Internal view of resting spore. (i) Girdle bands of the vegetative cell.

Table 5.3. Comparison of the morphological characteristics of *Chaetoceros socialis sensu stricto*, *C. gelidus*, *C. sporotruncatus* sp. nov. and *C. dichatoensis* sp. nov.

Type species	***Chaetoceros socialis sensu stricto***	***Chaetoceros gelidus***	***Chaetoceros sporotruncatus***	***Chaetoceros dichatoensis***
Colony	Curved chains joined to form more or less spherical colonies	Curved chains joined to form more or less spherical colonies	Curved chains joined to form more or less spherical colonies	Curved chains joined to form more or less spherical colonies
Aperture	Hexagonal, wide in pervalvar axis	Hexagonal, narrow in pervalvar axis	Hexagonal, wide in pervalvar axis	Hexagonal, wide in pervalvar axis
Intercalary setae: origin	More distant from the corner	Just within the corner	More distant from the corner	More distant from the corner

Type species	*Chaetoceros socialis sensu stricto*	*Chaetoceros gelidus*	*Chaetoceros sporotruncatus*	*Chaetoceros dichatoensis*
Basal part of the setae	Long	Short	Long	Long
Cross over	Chain margin	Chain margin	Chain margin	Chain margin
Short setae: morphology	Spirally arranged spines and poroids. Large solitary pores	Spirally arranged spines and poroids. Large solitary pores	Spirally arranged spines and poroids. Large solitary pores	Spirally arranged spines and poroids. Large solitary pores
Long seta: morphology	Spirally arranged spines and poroids; large solitary pores	Spirally arranged spines and poroids; large solitary pores. Spines only at end.	Spirally arranged spines and poroids; large solitary pores. Spines only at end.	Spirally arranged spines and poroids; large solitary pores. Spines only at end.
Valves	Elliptical to circular	Elliptical to circular	Elliptical to circular	Elliptical to circular
Annulus	Roundish central annulus	Roundish central annulus	Roundish central annulus	Roundish central annulus
Valve ornamentation	Costae radiating from the central annulus and converging to setae origin	Costae radiating from the central annulus and converging to setae origin	Costae radiating from the central annulus and converging to setae origin	Costae radiating from the central annulus and converging to setae origin
Rimoportula	Central	Central	Central	Central
Girdle bands	Transverse costae, several scattered minute poroids	Transverse costae, several scattered minute poroids	Transverse costae, several scattered minute poroids	Transverse costae, several scattered minute poroids
Spore: shape	Biconvex; hemispherical-conical valves	Biconvex. Mantle with crest made of fused projections.	Biconvex; valves as truncated cones.	Biconvex; hemispherical-conical valves
Spore: ornamentation	Spines (at times bifurcate); one row of spines on the primary valve margin	Smooth. Covered by crest on both valve	Primary valve with raised lenticular-shaped structures at the apex	Spines, often with an expanded base.
Apical axis (µm)	11.4±1.3 (8.9−13.3)	7.1±2.7 (3.6−16.9)	11.2±2.3 (7.2−15.0)	6.4±1.4 (3.9−10.0)
Pervalvar axis (µm)	8.6±2.0 (6.1−14.4)	9.7±1.8 (5.0−14.7)	11.3±4.13 (7.2−27.8)	11.6±2.7 (6.1−22.2)
Aperture in pervalvar axis (µm)	4.9±1.5 (2.8−8.9)	2.0±0.6 (0.7−3.3)	4.6±1.3 (2.2−6.7)	2.6±0.9 (1.1−6.7)
Poroids on the setae (N°/µm)	21±1 (n=3)	NA	20±1 (n=11)	21±1 (n=16)
Spines (setae N°/µm)	3	NA	3	3

***Chaetoceros radicans* Schütt**

Fig. 5.10. a–l

I have examined the morphology of five strains of *C. radicans* isolated from Las Cruces, Chile (Table 5.1). The gross morphology and ultrastructure of these strains matched the description provided by Cupp (1943), Hernandez-Becerril (1996), Shevchenko *et al.* (2006); Sunesen *et al.* (2008); Ishii *et al.* (2011) and Lee *et al.* (2014). Chains long, either straight, curved or slightly twisted (Figs 5.10. a–b). Cells are rectangular in girdle view, separated by a narrow hexagonal aperture (Figs 5.10. a–b); occasionally a broader aperture is observed (Fig. 5.10. d). Cells contain one large chloroplast. Delicate setae emerge inside the corners of the valves with a basal part and cross over at chain margin (Fig. 5.10 d). Setae diverge along the transapical axis (Fig. 5.10. c). Intercalary setae are differentiated into normal (Fig. 5.10. f) and specialised setae (Fig. 5.10. c). Specialised intercalary setae have capilli of variable length, at times branched (Fig. 5.10. c), while normal setae appear smooth in light microscopy.

The specialised and normal intercalary setae are circular in cross section and are ornamented with spines and rows of small poroids organised in a spiral pattern, along with large solitary pores (Figs 5.10. e–f). All strains stopped producing specialised setae after a few generations in culture. Terminal setae have similar ultrastructure as the normal intercalary setae, with a denser pattern of spines at the base of the setae (Fig. 5.10. e). The valve face is elliptical to circular, weakly silicified with a central annulus from which dichotomously branching costae radiate towards the valve margins (Fig. 5.10. h). The area within the annulus is hyaline (Fig. 5.10. h). A narrow hyaline rim was present on the edge of the valve surface (Fig. 5.10. h). Terminal valve with a central rimoportula, which is slit-like on the inner valve face and has the shape of a short flattened tube on the outer face (Figs 5.10. i–j). The rimoportula is situated inside the central annulus. Intercalary valves without rimoportula (Fig. 5.10. h).

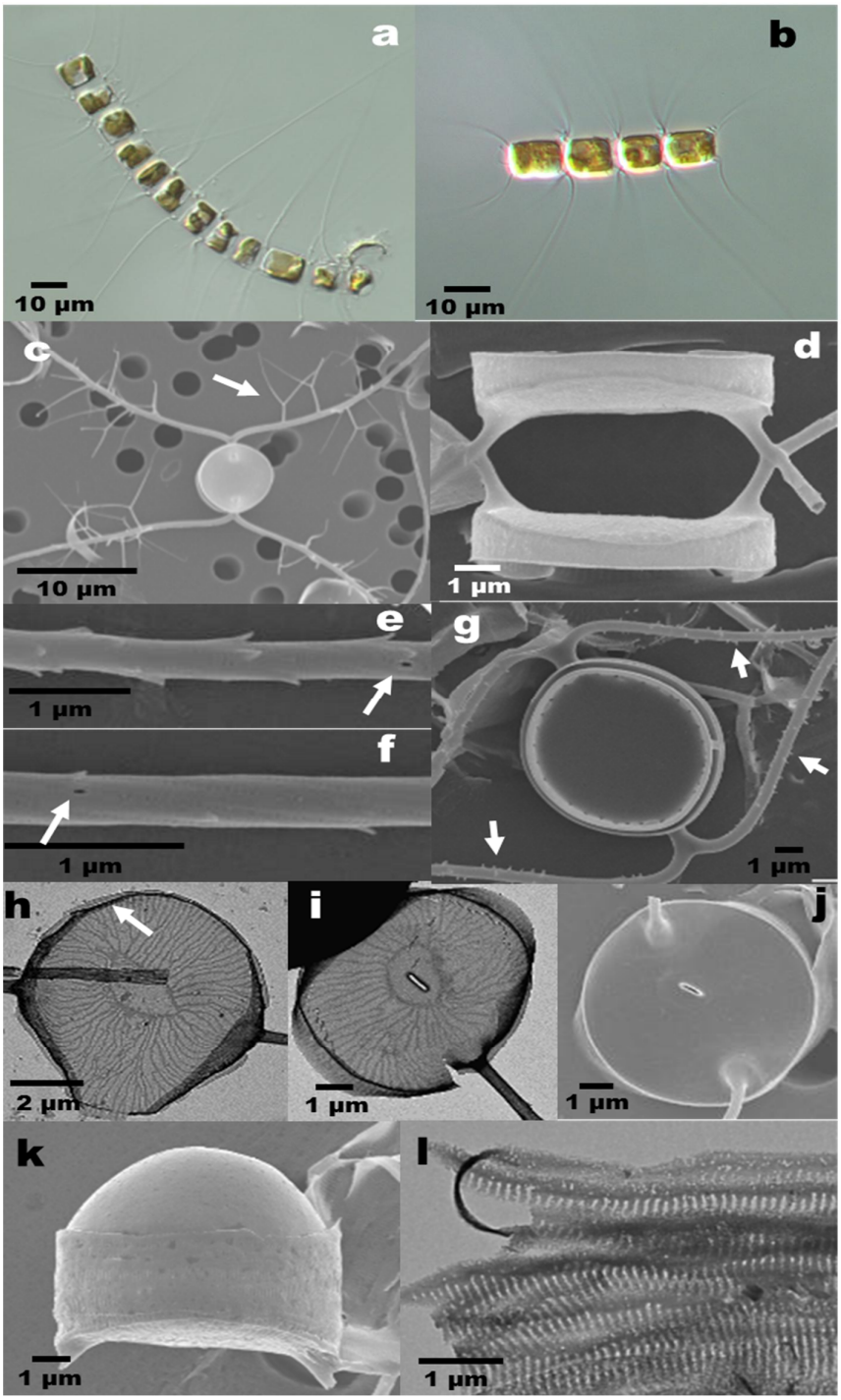


Fig. 5.10. *Chaetoceros radicans*. Strains Ch1B4 (b, d–i and k–l), Ch10A3 (c and j), Ch11A4 (a) from Las Cruces, Chile. LM (a–b), SEM (c–g and j–k), TEM (h–i and l). (a–b) Chains in girdle view. (c) Specialised inter-calary setae (arrow marks branched-capilli). (d) Aperture. (e) Basal part of a terminal seta; spirally arranged spines are visible and pore is arrowed. (f) Middle portion of an intercalary seta; the arrow marks a large pore. (g) The spore-holding basal plate containing the secondary valve of a spore; note the modified setae with short spines (arrowed); the line of puncta on the inner secondary valve are marked with arrowheads. (h) Intercalary valve with radially branched costae and a central annulus; arrow marks the hyaline valve rim. (i) Terminal valve with the central, slit-shaped rimoportula (internal view). (j) Terminal valve with the central, tube-shapes rimoportula (external view). (k) Smooth spore. (l) Girdle bands.

Girdle bands with numerous parallel costae with minute poroids (Fig. 5.10. l). The apical axis is 9.3±2.1 (5.4−12.8) µm long, the pervalvar axis 10.1±3.5 (4.3−16.7) µm long, and the length of the aperture in the pervalvar axis is 2.0±0.7 (1.1−3.9) µm. Resting spores are plano-convex: the primary valve is dome-shaped and the secondary valve is flat; both valves have a smooth surface (Fig. 5.10. k). Resting spores are present in pairs and adhere with their flat valve to a basal plate. The setae of the basal plate have a distinct orientation: they fuse slightly outside the chain margin, then separate, turning around the cell (Fig. 5.10. g). The setae emerging from the basal plate are ornamented with spirally arranged small spines and organised with rows of poroids and scattered pores (Fig. 5.10. g).

***Chaetoceros cinctus* Gran**

Fig. 5.11. a–j

I have examined the morphology of seven strains of *C. cinctus* isolated from Las Cruces, Chile (Table 5.1). The gross morphology and ultrastructure of these strains matched the description provided by Cupp (1943), Shevchenko *et al.* (2006). Chains long, either straight, curved or slightly twisted (Figs 5.11. a–b). In girdle view, cells appear rectangular. Cells with narrow hexagonal apertures (Fig. 5.11. c). Each cell contains a single large chloroplast. Delicate setae emerge inside the corners of the valves with a basal part and cross over at chain margin (Fig. 5.11. c). Setae diverge along the transapical axis. Setae are circular in cross section and are covered with spirally arranged spines and poroids (Fig. 5.11. d–e), along with large solitary pores (Fig. 5.11. e). Valve face elliptical to circular, weakly silicified with a central annulus from which dichotomously branched costae radiate towards the valve margins (Fig. 5.11. f). A small hyaline rim was present on the edge of the valve surface (Fig. 5.11. f). The terminal valve has a central rimoportula, which is slit-like on the inner valvar face (Fig. 5.11. h) and has the shape of a tube on the outer valvar face (Fig. 5.11. g).

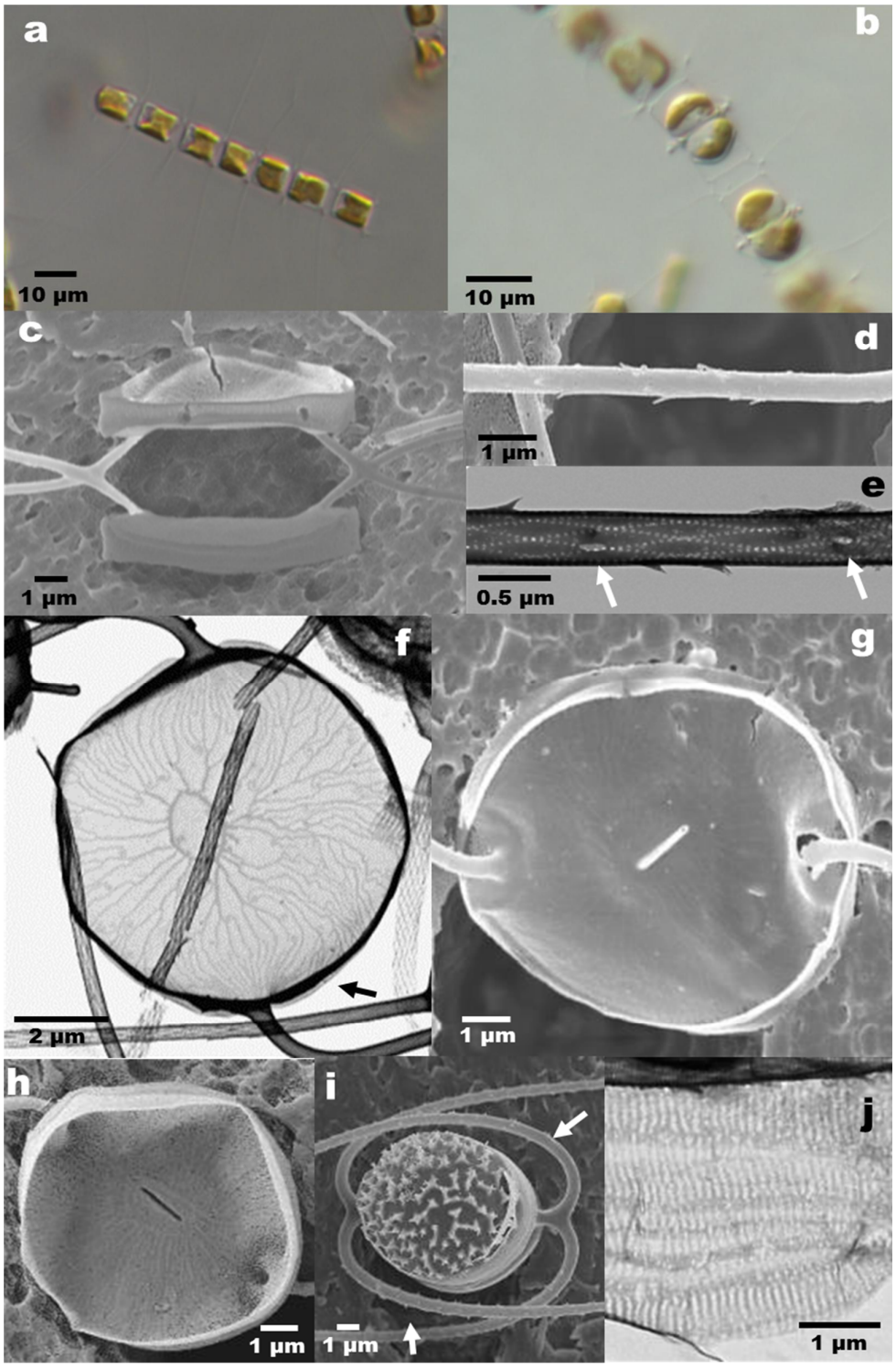


Fig. 5.11. *Chaetoceros cinctus*. Strain Ch10B3 from Las Cruces, Chile, (a–j). LM (a–b), SEM (c–d and g–i), TEM (e–f and j). (a) A chain in girdle view. (b) Paired resting spores in the chain. (c) Aperture. (d) Intercalary seta on which spirally arranged spines are visible. (e) Terminal seta ornamented with spirally arranged spines and poroids, and large pores (arrowed). (f) Intercalary valve with radially branched costae and the central annulus; the arrow marks the hyaline rim. (g) External view of a terminal valve with a central tube-like rimoportula. (h) Internal view of a terminal valve with a central slit-shaped rimoportula. (i) Spore located on the basal plate; the arrow indicates the small spines on the setae. (j) Girdle bands.

Girdle bands are very delicate and ornamented with numerous parallel costae and poroids (Fig. 5.11. j). The apical axis is 8.7±2.2 (3.3−16.7) µm long, the pervalvar axis 9.6±3.1 (3.3−16.7) µm long, and the length of the aperture in the pervalvar axis is 2.3±0.9 (1.0−3.4) µm. Resting spores are organized in pairs and have a plano-convex shape (Fig. 5.11. b). The surface of the primary valve is covered with numerous spines, often with an expanded base that connects adjacent spines (Fig. 5.11. i). The spores adhere in pairs to a basal plate holding the adjacent spores together (Fig. 5.11. j). The setae of the basal plate have a distinct orientation: they fuse slightly outside the chain margin than separate, turning around the cell. The setae of the basal plate are ornamented with spirally arranged small spines (Fig. 5.11. j).

Discussion

The results of the present study show that strains identifiable in light microscopy as *C. socialis*, isolated from southern and western Europe (the Gulf of Naples, Italy, and Roscoff France) differ genetically from those off the central Chilean coast. In addition, those gathered from Chile belong to two species new to science, here described as *C. sporotruncatus* and *C. dichatoensis*. Moreover, the results demonstrate that Chilean strains identified as *C. radicans* and *C. cinctus* are morphologically distinct, though phylogenetically closely related.

Phylogeny of the *Chaetoceros socialis* complex: the *Chaetoceros socialis* complex is monophyletic according to both the 18S and 28S rDNA phylogeny. This clade comprises four terminal clades representing four species: *C. socialis sensu stricto*, *C. gelidus*, and the two species new to science, *C. sporotruncatus* and *C. dichatoensis*. The clade with *C. sporotruncatus* contains also strain CCMP172. The morphological description of this strain was provided by Chamnansinp *et al*. (2013), but they did not describe it as a different species because this strain was the only one available to them, and it did not produce resting spores. Resting spore information is available for strains from the Chilean central coast. and strain

CCMP172 is now assigned to *C. sporotruncatus*. The strains of *C. dichatoensis* forms a sister clade to *C. sporotruncatus*.

Genetic identity of the 18S and partial 28S rDNA sequences of all strains from the GoN and Roscoff suggest that all these strains belong to a single species, *C. socialis* (sensu stricto). This European group of strains differs from South China Sea strain YL1 in a transition A↔G at position 469 in the partial 28S sequence alignment, and from the Thai strain No_1 at five positions near the 5'-end of the sequence alignment. It cannot be excluded that these strains represent a different population/species, but this needs to be confirmed by obtaining multiple sequences from those regions. In the other three species, i.e., *C. gelidus*, *C. sporotruncatus* and *C. dichatoensis*, no variation was detected among the sequences used in the study.

Morphological comparison: based on the molecular phylogeny four species were revealed, whereas all the species showed similar vegetative morphology in LM. The morphological characters of vegetative cells and spores of the different species within the *C. socialis* complex are provided in Table 5.3. The similar morphology of the vegetative cells of the species in this complex makes their differentiation and correct identification in light microscopy very difficult or impossible since they all share the same primary and secondary structure of the colony, the orientation pattern of the setae, and valve ultrastructure.

However, there are a few characters of the vegetative cell that can be helpful in distinguishing *C. gelidus* from the other members of the complex. This is the case of the aperture size, which is narrower in transapical axis in *C. gelidus*, and of the location from which the setae emerge, which is closer to the margin of the valve face in *C. gelidus.* An ultrastructural character that can help to differentiate *C. socialis sensu stricto* from the other species in the complex is the presence of spines along the whole of the long seta, whereas the other species possess smooth setae with spines present only in the distal part. The arrangement/number of spines and pores on the setae has been used to differentiate some

Chaetoceros species. For example, species with very similar morphology, such as *C. contortus* and *C. compressus* or *C. decipiens* and *C. lorenzianus*, can be differentiated using these ultrastructural details (Lee & Lee 2011; Lee *et al.* 2014).

Resting spore morphology is the most prominent character to differentiate among species of the *C. socialis* complex (see Table 5.3). The spores of *C. socialis sensu stricto* are ornamented with spines, at times with an expanded base, and have a ring of spines on the margin of the primary valve. Spores of *C. gelidus* have smooth valves and the mantle of both valves have a distinctive crest of fused projections. In *C. sporotruncatus*, the spores are biconvex with truncated valves and the primary one is ornamented with raised lenticular-shaped structures. Finally, the spores of *C. dichatoensis* are ornamented with spines with an expanded base that form ribs connecting several spines.

In the literature, there are reports and illustration of spores attributed to *C. socialis*. Hargraves (1979) described spores covered by spines, with a row of spines at the margin between the valves and the mantle; the author attributed this spore to *C. sociale* var. *radians* (Figs 42−47, in Hargraves 1979). These spores match with the description of the spores of *C. socialis sensu stricto.* Another spore morphotype with smooth valves and surrounded by a perforated collar was attributed to *C. sociale* var. *sociale* (Figs 48−50, in Hargraves 1979). This spore morphotype corresponds to the spore of *C. gelidus.* Two more morphotypes of spores - tentatively attributed to *C. socialis* - were described by Hargraves (1979). One presented a collar, and the valves were ornamented by circular or ovoid pores with a thick rim; some of these pores penetrated to the spore interior (Figs 52−54, in Hargraves 1979). Another morphotype, collected along the coast of Peru, lacked the collar and presented 'raised slots' on the top of one valve (Figs 55−56, in Hargraves 1979). This latter spore morphotype corresponds to the spore of *C. sporotruncatus*. A spore similar to that of *C. sporotruncatus*

was reported also from the Mexican coast on the Pacific Ocean of by Hernández-Becerril (1996; see Plate 51, 4−5).

Can spore morphology reveal species diversity? Is it important to know it? It is known that resting spore morphology can be helpful in differentiating among the species of the complex when considered together with the morphological characters of the vegetative cells, it has to be noted that some of spores resemble those produced by different *Chaetoceros* species. This represents a problem for the correct identification of spores at the species level, e.g. when they are recorded alone in sediment traps or in sediment cores, unless spores are observed in SEM that allows detecting minor differences. The spores of *C. socialis sensu stricto* at times resemble those of *C. dichatoensis*; however the first species presents a distinctive line of spines on the margin of primary valve. Another example is the spore of *C. gelidus*, which somewhat resembles the spore of *C. contortus* (see figs 27−29, in Kooistra *et al.* 2010) and *C. protuberans* (see figs 91−94, in Kooistra *et al.* 2010). Nevertheless, the spore of *C. gelidus* is surrounded by crests on both valves, whereas those of *C. contortus* and *C. protuberans* show crest only on the primary valve (Kooistra *et al.* 2010). The resting spore of *C. curvisetus* and *C. pseudocurvisetus* show a crest on both valves as in *C. gelidus*, but the fissures are fine and the crests have a fold at the edge towards the inside (see Kooistra *et al.* 2010, for *C. curvisetus* spores figs 44−45).

Taxonomy and phylogeny of *C. radicans* and *C. cinctus*

From the phylogenetic analyses, it was inferred (with high bootstrap support) that *C. radicans* and *C. cinctus* (which are sister species) are sister to the *C. socialis* complex. *Chaetoceros radicans* was described from the Atlantic Ocean (only this very general information was provided) by Schütt (1895). Further studies examined the ultrastructure of this species in EM (Hernández-Becerril 1996; Sunesen *et al.* 2008; Lee *et al.* 2014) and the Chilean strains examined in the present study matched these descriptions with only minor

differences. Specimens obtained from field material from the Atlantic Ocean along the Argentinian coasts showed smooth setae on the basal jacket holding the pair of spores (fig. 12 C and F, in Sunesen *et al.* 2008), while the strains isolated from Chilean waters have the setae of the basal plate ornamented with numerous short spines. The spores of *C. radicans* described in the present study have a more vaulted primary valve as compared to the Atlantic specimens (Sunesen *et al.* 2008, figs 12. D-E). The strain WK-140306 isolated from the West Sea of Korea and identified as *C. radicans* showed smooth normal intercalary setae without pores (Lee *et al.* 2014, figs 113−115, 134), while in the Chilean strains, the normal intercalary setae were ornamented with rows of poroids and with few large solitary pores. Unfortunately, molecular sequences were not available for both the Argentinian and the South Korean strains.

The clade of *C. radicans* contained one branch constituted by strain CCMP197 from the North Sea. Unfortunately, no morphological information is available for this strain. Further detailed studies examining ultrastructure and molecular signatures of other strains collected over a broader geographic range and studying the level of plasticity of morphological characters in cultured strains and in natural populations will help better defining the identity of this species.

Chaetoceros cinctus can be easily differentiated from *C. radicans* as it lacks the characteristic branched intercalary setae. Moreover, the resting spores of *C. cinctus* are ornamented with small spines, while those of *C. radicans* are smooth.

Conclusions and outlook

In the present chapter, I investigated the morphological and molecular diversity within the *Chaetoceros socialis* complex, where I describe the morphology and ultrastructural details of two new species *C. sporotruncatus* sp. nov., and *C. dichatoensis* sp. nov., that share

similar morphology as that of *C. socialis sensu stricto*. Phenotypic and genetic evidence suggests the presence of multiple species within the *C. socialis* complex, which share similar morphologies. Some of the members of these taxa show allopatry, as in the case of *C. gelidus,* which is restricted to cold waters, whereas *C. socialis sensu stricto* is present in warm waters. In addition, members of this complex show sympatry, i.e., in the case of *C. sporotruncatus* and *C. dichatoensis*. All these species have almost indistinguishable vegetative morphologies, and can only be differentiated using spore morphology. It is evident that the *C. cinctus* and *C. radicans* are morphologically and genetically distinct, and share a relationship with the *C. socialis* complex.

Characterization of introns in the nuclear encoded ribosomal RNA of the planktonic diatom family Chaetocerotaceae

Abstract

Introns are known to occur all over the tree of life; in archaea, bacteria, and eukaryotes, and even in viruses. Introns are classified in four major groups: spliceosomal introns, tRNA introns, group I introns and group II introns. Whilst gathering 28S and 18S rDNA sequences of the various species in the family Chaetocerotaceae, two types of introns were encountered, namely long ones (ca. 400 bp) and short ones (ca. 100 bp). They were encountered at 14 positions in the 18S rDNA alignment and one in the partial 28S rDNA alignment. Secondary structure analyses of the long introns revealed sequence conservation within the regions known to be functionally important for in vitro self-splicing of group-I introns. To better understand the phylogenetic relationships of the group I introns, which were detected in the 18S rDNA, an alignment was generated, including our Chaetocerotacean intron sequences and ones from members of other algal families, gathered from GenBank. Phylogenetic analysis of this alignment revealed that most of the group I introns formed a distinct lineage, defined by insertion sites within the 18S rDNA. The short introns, the vast majority of introns encountered, were scattered all over the 18S rDNA sequences and at the one position in the partial 28S rDNA sequences. These short introns showed the conserved signature regions typical for spliceosomal introns and highly variable regions in between, though with characteristic base composition. Alignment of these variable regions was next to impossible, and as a consequence phylogenetic relationships remained unresolved. Only closely related species shared similar introns at the same locations, and conspecific strains from the same location tended to share highly similar or identical introns. The phylogenetic relationship among these introns did not corroborate with that of the phylogeny of Chaetocerotaceae species inferred from the SSU or LSU core sequences in which the introns were encountered suggesting that they are transferred horizontally among different species; or even across lineages in the tree of life as some of the Chaetocerotacean introns even clustered with red algal lineages. The introns were encountered only at the conserved locations in the rDNA genes. The introns were not distributed all over the Chaetocerotacean diversity but instead concentrated in three clades. Yet, not all the members of these clades contained introns. The strains containing introns were mostly encountered during warm condition, a possibility for resistance/infection for parasites. Or this may even hint at the existence of different populations in different seasons at the Gulf of Naples; as in *C. vixvisibilis*, an intron was present in all of the summer strains and absent in strains collected during other seasons.

Introduction

The Chaetocerotaceae constitutes a family of centric diatoms in the marine plankton. It includes two genera: *Chaetoceros* and *Bacteriastrum*. The family is highly diverse (Guiry & Guiry 2016). In a series of taxonomic studies exploring the diversity of the family, a series of authors (Kooistra *et al.* 2010; Chamnansinp *et al.* 2013, 2015; Li *et al.* 2013, 2015; Bosak *et al.* 2015) have used a combination of morphological and molecular information to delineate species. All these authors used a part of the nuclear-encoded large subunit ribosomal RNA region (28S rDNA) as a marker of choice, rather than the nuclear-encoded small subunit ribosomal RNA region (18S rDNA), which is surprising because this marker is commonly used in phylogenetic studies of the diatoms. In addition, 18S rDNA sequences of Chaetocerotaceae are not well represented in GenBank, in spite of the ecological importance of the family.

In order to assess the morphological and genetic diversity of the family Chaetocerotaceae, I sequenced the partial 28S rDNA from a large series of strains and aimed at gathering the full-length 18S rDNA sequences of all these strains as well. Initially, I failed to PCR-amplify the 18S sequences of several of these strains and it soon became clear that failure was taxonomy-related. Using internal primers, normally deployed for sequencing, as PCR primers, I managed to amplify the 18S of these strains in a series of overlapping parts to assure sequence identity, and then pieced these parts together. While doing so, I noticed the presence of one to several introns in the rDNA sequences of many of the strains. In the present chapter, I focus on these introns.

Introns were discovered in the protein-coding genes of adenovirus independently by Berget *et al.* (1977) and Chow *et al.* (1977) and identified subsequently also in transfer RNA- (tRNA) and ribosomal RNA (rRNA) genes. For this discovery, Phillip A. Sharp and Richard J. Roberts shared the Nobel Prize in Physiology or Medicine in 1993. The term intron was introduced by Walter Gilbert (Gilbert 1978) and is derived from "intragenic region," i.e., a region within a gene. An intron is a portion of a sequence that does not code for anything; it is transcribed, but subsequently excised from the transcripts by an RNA splicing mechanism during RNA maturation, and in the case of protein-coding RNA (mRNA), preceding translation. Therefore, "intron" refers to the DNA sequence within a gene as well as to the RNA sequence in the transcript. Introns are now known to occur in a range of genes, including protein-coding genes and ribosomal RNA (rRNA) and transfer RNA (tRNA) coding regions. They can be found in all lineages in the tree of life, including bacteria,

archaea and eukaryotes, as well as in viruses (Michel & Westhof 1990; Lambowitz & Belfort 1993; Saldanha *et al.* 1993; Yamada *et al.* 1994; Bhattacharya *et al.* 1996a; Müller *et al.* 2001; Haugen *et al.* 2005).

Four distinct classes of introns are now recognized (Alberts 2008):

- Spliceosomal introns: introns in nuclear protein-coding genes that are removed by spliceosomes;
- Transfer-RNA (tRNA) introns: Introns in nuclear and archaeal tRNA genes that are removed by proteins;
- Self-splicing group I introns that are removed by RNA catalysis.
- Self-splicing group II introns that are removed by RNA catalysis

Spliceosomal introns are present in the nuclear protein-encoding genes and are processed and spliced by spliceosomes. These introns are characterized by their short lengths (49-199 bp). They are found usually in messenger RNA (mRNA), but, for instance, in fungi, they are observed also in the rRNA genes (Guthrie & Patterson 1988; Rogers *et al.* 1993; Bhattacharya *et al.* 2000). Spliceosomal introns are defined by conserved sequences located at the boundaries between introns and exons (Padgett *et al.* 1986). At the 5' end of the intron (the donor site), there is an almost invariant sequence GU within a less conserved region. The 3' end of the intron (the splice acceptor site) exhibits a virtually invariant AG sequence. The conserved 5'-ends and 3'-ends are recognized by spliceosomal RNA molecules that than in their turn initiate the splicing reactions (Guthrie *et al.* 1988). Upstream (in the 5'-direction) from the AG there is a region high in pyrimidines (C and U). Further upstream from this polypyrimidine tract is the branch point, which includes an A involved in the bridge formation. It becomes covalently linked to the 5' end of the intron during the splicing process, generating a branched intron. The consensus sequence for a spliceosomal intron is expected to be: GG--GURAGU (donor site) ... intron sequence ... YURAC (branch sequence 20-50 nucleotides upstream of acceptor site) ... Y-rich-NCAG--G (acceptor site) (Black 2003; Clancy 2008). Apart from these conserved regions, there is no apparent pattern of sequence conservation (Bhattacharya *et al.* 2000).

tRNA introns occur in the tRNA genes of eukaryotes and are known to interrupt the anticodon loop (one base 3' to the anticodon) (Tocchini-Valentini *et al.* 2011). tRNA introns are also common in tRNA genes of Euryarchaeota (extremophiles) and in tRNA- and rRNA-genes of the Crenarchaeotes (Marine Archaea; Lykke-Andersen *et al.* 1997). tRNAs are

found in nuclear and archaeal transfer RNA genes and are excised from the RNAs by proteins.

Group I and group II introns are present in mRNA-, tRNA- and rRNA-coding genes. Upon transcription they undergo extensive interactions, leading to complex three-dimensional secondary structures, which in their turn initiate self-splicing. These introns are distinguished on the basis of their conserved regions and their folding patterns. Group II introns are a type of retroelement found in bacterial and eukaryotic organellar genomes and are considered to be the ancestors of spliceosomal introns (Michel & Lang 1985; Dai & Zimmerly 2002; Khan & Archibald 2008). A fifth type, group III introns, are proposed, but it still does not have sufficient biochemical support or information (Copertino & Hallick 1993). Group III introns are also known as twintrons, i.e., introns-within-introns, found in the Euglenoid chloroplast DNA. They are characterized by a relatively short and uniform fragment length of ca. 100 bp and their RNAs are generally U-rich (Copertino & Hallick 1993).

Group I and group II introns are widespread in different lineages, but distributed irregularly (Dujon 1989). These introns are found in fungi and algae (Dujon 1989; Bhattacharya *et al.* 1996b; Johansen *et al.* 1997; Hafez *et al.* 2012), in bacteriophages and eubacteria (Lambowitz & Belfort 1993), and in eukaryotic viruses (Yamada *et al.* 1994). They abound in the mitochondrial and chloroplast genome, including in their protein-coding genes. In the nuclear genome, they seem restricted to the rRNA genes. These introns seem to be transferred horizontally between different evolutionary lineages (Vader *et al.* 1994; Yamada *et al.* 1994; Belfort & Perlman 1995).

In eukaryotic nuclear genomes, group I introns are mostly found in the rRNA genes in which they show remarkable variation in length, from 56 bp (Grube *et al.* 1996) to 3000+ bp (Lambowitz & Belfort 1993), yet all share a distinct, conserved secondary structure (Cech 1986; Doudna & Cech 2002). This consists of a series of paired regions (P1 to P10) that assemble to form a catalytic ribosome (Vicens & Cech 2006). The paired regions exhibit three domains: a substrate domain (P1-P2-P10), a scaffold domain (P4-P5-P6) and a catalytic domain (P3-P7-P8-P9), connected by non-Watson-Crick bonds (Kim & Cech 1987; Michel & Westhof 1990; Woodson 2005; Golden 2008). Group I introns are classified into at least five main groups (1A, 1B, 1C, 1D, and 1E) which in turn are divided into ten subgroups (IA1, IA2, IA3, IB1, IB2, IB3, IB4, IC1, IC2, and IC3). This classification is based on their sequence motifs, location, and characteristic structural features in the 5'- and the 3'-ends (Einvik *et al.* 1998; Li & Zhang 2005; Mitra *et al.* 2011). Also, the insertion positions of

group 1 introns in the 18S rRNA gene seem to be conserved (Bhattacharya *et al.* 1996a, 1996b). With the availability of more introns, new positions for introns have been revealed (Gargas *et al.* 1995; Müller *et al.* 2001).

Group I introns are capable of self-splicing (Cech 1986; Doudna & Cech 2002) by a series of guanosine initiated trans-esterification reactions (Foley *et al.* 2000). The P1 (5'-end) and P10 (3'-end) regions possess the splicing sites (Vicens & Cech 2006), which are formed by base pairing to the internal guide sequence (IGS), which often forms triple bases involving P1, IGS, and P10. The primary structure (sequence) of the group I introns varies, except for few crucial nucleotides at the active sites, although the core secondary and tertiary structures are conserved (Michel & Westhof 1990). The conserved core structure includes the P, Q, R, and S sites. These sites occur always in the same order i.e. 5'-P-Q-R-S-3' along the intron, but their length may vary from a few to a hundred nucleotides (Cech 1988). Some group 1 introns also act as mobile genetic elements that jump within genes (Johansen *et al.* 1997). Insertion of a particular intron at the same position in homologous genes is called homing (Dujon *et al.* 1989) and is catalysed by an intron-encoded DNA endonuclease (Belfort & Roberts 1997).

In group II introns, splicing occurs in two transesterification steps. Group II introns contain six domains (D1-D6). Domain 1 (D1) is known to possess catalytic activity and is also the largest non-coding domain. The D2 function is unclear, but it is thought to play a role in having tertiary interaction with D6 and D1 (Chanfreau & Jacquier 1993, 1996). D3 seems to play a role in splicing, as D3 deletion has impaired in in vitro catalysis (Podar *et al.* 1995). The function of D4 is not clear. Group II introns possess a unique conserved region containing 5'-GYGYG and 3'-AY.

In my study, I gathered all the introns detected in the 18S and partial 28S rDNA sequences gathered from my Chaetocerotacean strains and aligned these with similar introns gathered from the nucleotide collection in GenBank using BLAST. I acquired the alignment and secondary structure of introns from the Comparative RNA Web (CRW) site operated by the Gutell Lab, University of Texas, Austin, USA (http://www.rna.ccbb.utexas.edu; Cannone *et al.* 2002) for structural analysis. A phylogenetic approach was used to investigate the evolutionary relationship among the obtained introns. Secondary structure models of introns were inferred using those of closely related introns in this database as guidance.

Materials and Methods

Taxa studied: in the present study, inserts/introns were studied in the 18S rDNA and partial 28S rDNA regions in *Chaetoceros* and *Bacteriastrum* species. In the first instance, the introns in my study were characterized based on their length and internal structures. For convenience, the gene regions present in my alignment are indicated by red horizontal lines, and the introns within the genes with blue lines (see Figs 6.1.1–6.1.2).

Characterization of Intron: introns were characterized using their conserved sites. I tried to locate these sites, but it was not feasible for most of the introns because of length variability. I focused on the long inserts, as they were less common (three groups) than the small inserts and they showed several conserved regions and few variable regions. I tried to locate the conserved P, Q, R, and S sites in all the different form of the long inserts. I used a phylogenetic approach to detect closely related intron types, and then find possible information related to the type of intron.

Sequence analysis: for this study, I retrieved all the intron sequences from GenBank using BLAST search. All sequences resulting from the BLAST hits were downloaded irrespective of the hit percentage, but with length coverage above 60 % as a criterion for the selection of related sequences. In addition, I screened the following intron databases:

Bacterial Group II introns: http://webapps2.ucalgary.ca/~groupii/

Ares Lab Yeast Intron Database 4.1: http://intron.ucsc.edu/yeast4.1/

GISSD: http://www.rna.whu.edu.cn/gissd/index.html

CRW Site and project: http://www.rna.icmb.utexas.edu/DAT/3C/Alignment/

Upon retrieval of these sequences, I generated a local intron dataset with sequences belonging to different lineages, including eukaryote, fungi, bacteria, and viruses. In order to gather the best possible results, I did a BLAST search for my intron sequences as a query against this local dataset containing introns (see Chapter 2 for details on local dataset generation and BLAST analysis). The intron characterisation was done using a phylogenetic approach to place the introns into groups of closely related sequences, followed by and secondary structure analysis of a representative of each of these groups.

Phylogenetic inference: the most closely related hits of BLAST results were aligned using MAFFT v7.245 (Katho *et al.* 2002), and the resulting alignment further refined using SeaView v4.5.4 (Gouy *et al.* 2010). The obtained alignment was used to infer a phylogeny

of each of the uncovered groups using Neighbour Joining of pairwise dissimilarities (NJ). Bootstrap support for internodes was obtained using 1000 bootstrap replicates.

Secondary structure analysis: I downloaded from the CRW site secondary structures of reference introns that were recovered in the NJ-tree close to Chaetocerotacean introns and mapped the latter sequences manually over the downloaded secondary structure of the reference sequences. To characterise the functional elements in the Chaetocerotacean introns, I mapped the conserved and variable regions indicated in the reference sequence of the model species *Aureoumbra lagunensis* onto the alignment of my intron sequences. The secondary structures of Chaetocerotacean introns were produced by comparison against the proposed secondary structure of group I introns of Cech *et al.* (1994). This file was later edited to a bpseq format to obtain an automated secondary structure using RNApdbee (Antczak *et al.* 2014). Bpseq format was needed for tertiary folding, and this helps in reorganizing the secondary structure of the intron using the thermodynamic principle of entropy.

Results

Taxa and their introns: short introns were observed in the partial 28S rDNA sequences of *Chaetoceros* 'brush' *C. diversus* (only one strain), *C.* 'verylongsetae' (Na13C2), *Chaetoceros* sp., and *C.* cf. *pseudocrinitus* at only one position (see Fig. 6.1.1). In the 18S rDNA, introns were observed in *Bacteriastrum hyalinum, C. anastomosans, C.* 'brush'*, C. circinalis, C. decipiens, C. diversus, C. pseudocrinitus, C.* 'singlecells'*, C.* 'verylongsetae' and *C. vixvisibilis* whereas long introns were present in *B.* cf. *furcatum, B. jadranum, C. brush, C. decipiens, C.* cf. *pseudocrinitus, C. rotosporus, C.* 'singlecells' and *C. seiracantus*. Introns were found among 14 Chaetocerotacean species (see Fig. 6.1.2). The size of the long introns varied in length between 360 and 505 bp, and the small introns between 90 and 162 bp. In *C. vixvisibilis* and in *C.* 'singlecells', a small intron was detected inside another small intron. The long introns were further classified into three types based on their alignment patterns, whereas the small introns were variable in size and alignment patterns. The distribution of the introns was not uniform based on the 18S and partial 28S rDNA phylogenies (see Chapter 3), but the introns were present in most of the strains collected from the same geographic area, i.e., the Gulf of Naples, Italy. Notably, they were not always present in all of the strains of a species. For instance, strains of *C. vixvisibilis* collected beginning October 2014 possessed an insert in their 18S rDNA whereas those collected beginning July of that same year lacked this insert.

Sequence analysis of the long introns: from the BLAST analysis, I collected all the sequences that matched with my sequence of interest in GenBank. The length criterion of ≥60% coverage restricted the number of related sequences obtained. Also, from the local dataset assessment, I retrieved only a few sequences that were close to my marker of interest and these intron sequences were gathered together. A total of 62 sequences were returned from the BLAST analysis, all of them grouping with the long introns.

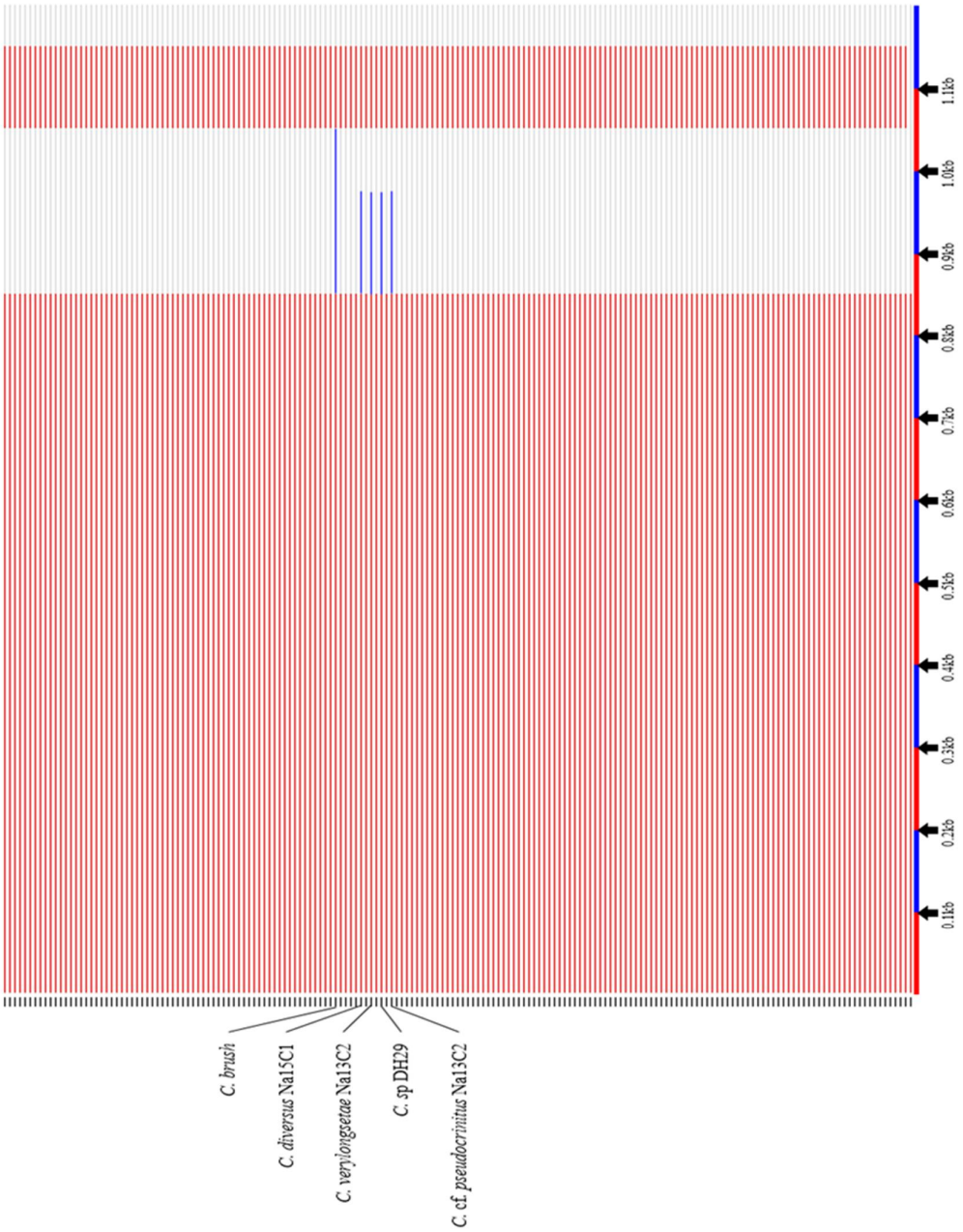


Fig. 6.1.1. The partial 28S rDNA alignment of the Chaetocerotacean species. The partial 28S rDNA sequence is ca. 800 bp. In the case of the Chaetocerotacean species, the alignment extends to ca. 1.1 kb. The red lines indicate the actual 28S sequences of different individuals. The blue line indicates the presence of inserts of variable length. Introns are present in only one position.

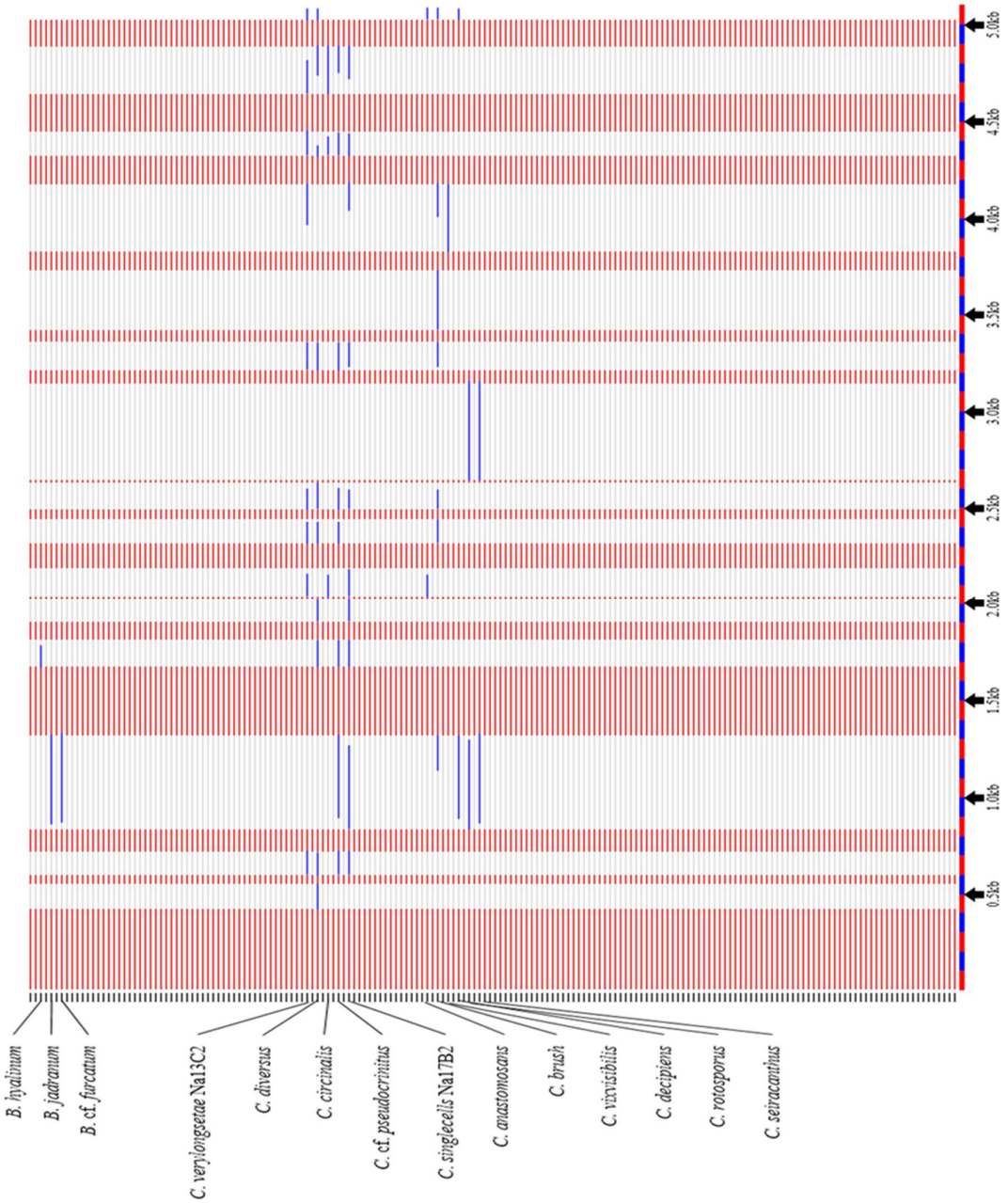


Fig. 6.1.2. The 18S rDNA alignment of the Chaetocerotacean species. The 18S rDNA sequence is ca. 1800 bp. In the case of the Chaetocerotacean species, the alignment extends to ca. 5 kb. The red lines indicate the actual 18S sequences of different individuals. The blue lines indicate the presence of inserts of variable length. A total of 15 positions are affected by introns in this alignment.

Phylogeny of the long introns: in the NJ tree inferred from the long introns (Fig. 6.3) the sequences grouped in two clades, the first of which (94%) contained the introns (at position 9 in Fig. 6.1.2) of *C. seiracanthus* and *C. rotosporus*, which grouped together with the sequence of Cps.S1046. Sister to this clade was a group of introns, including those from *Desmodesmus pannonicus*, *Phlyctochytrium planicorne* and *Selenastrum capricornutum* (all Chlorophyta)

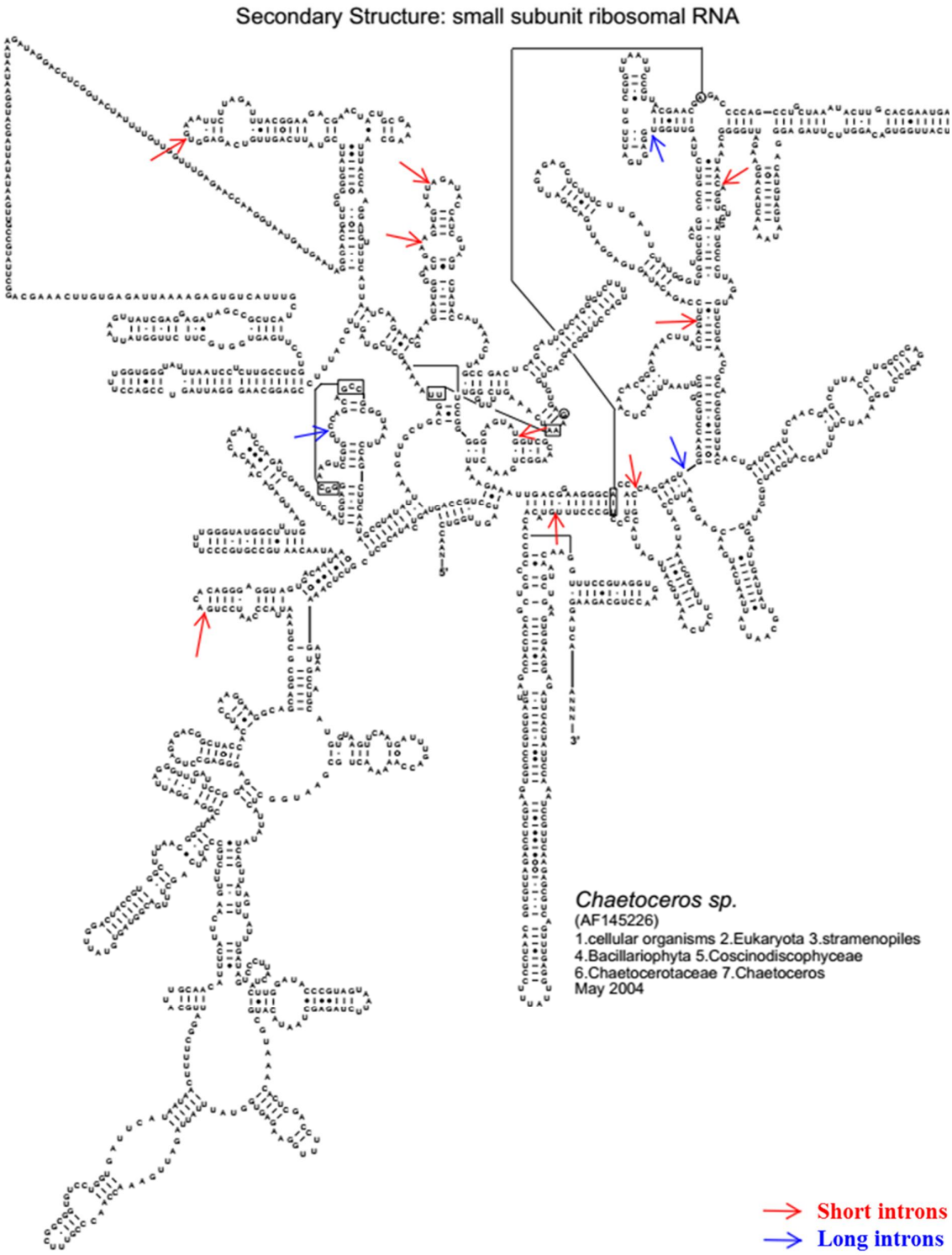


Fig. 6.2. *Chaetoceros tenuissimus* (AF145226) 18S rRNA secondary structure depicting the position of different introns in the alignment of Chaetocerotaceae. Note that the actual 18S rDNA sequences of *C. tenuissimus* strains did not contain any inserts. The rRNA secondary structure of this species is used merely as a map to show the location of inserts in the rDNA of various other Chaetocerotacean species. The red arrows indicate short introns (90-162 bp) and the blue arrows long introns (360-505 bp). The secondary structure was retrieved from www.rna.icmb.utexas.edu/SIM/4D/Diatom_nSSU/.

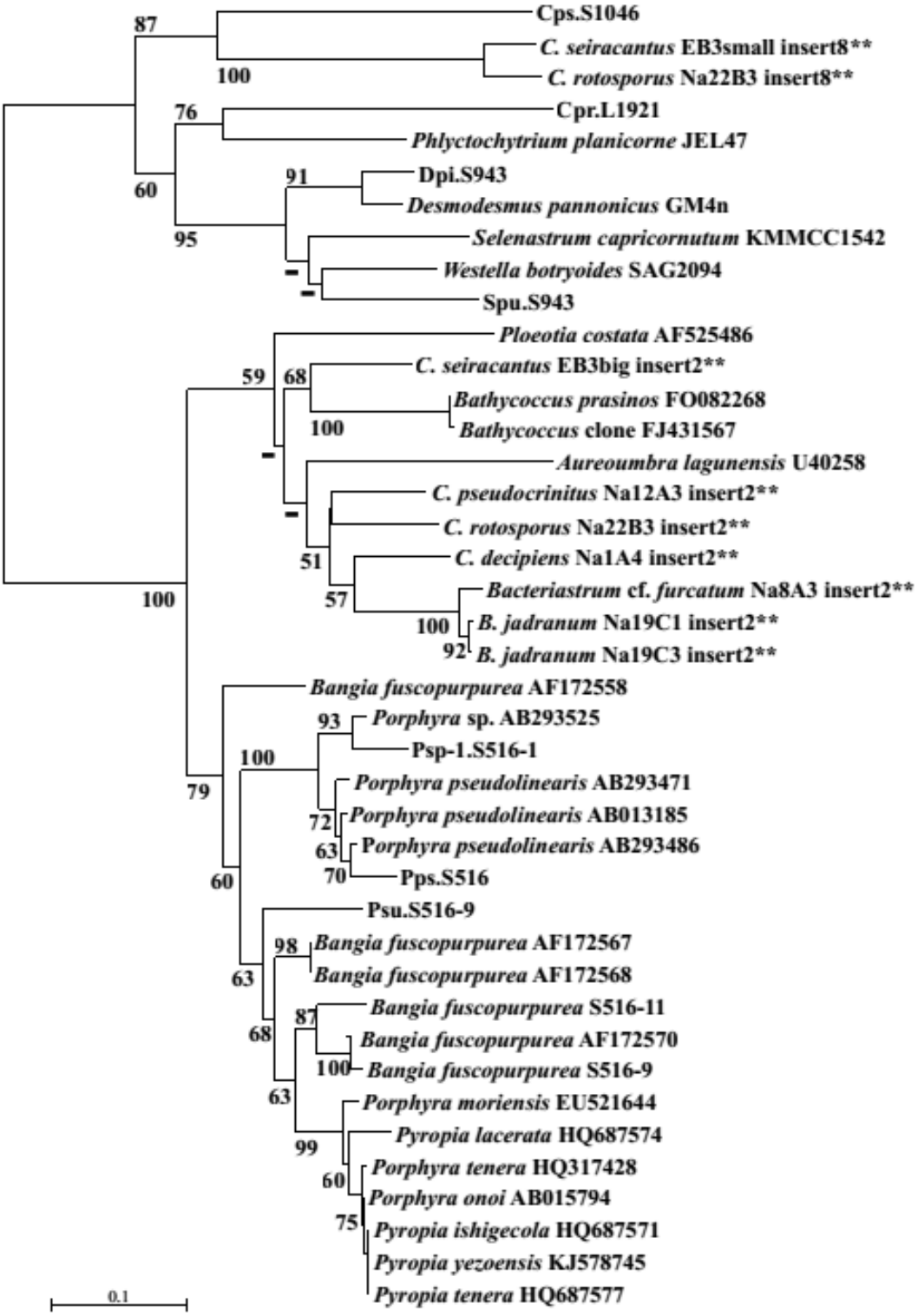


Fig. 6.3. Phylogeny of group I introns interrupting the nuclear SSU rDNA of Chaetocerotacean species. This tree was built with Neighbour Joining, using a K2P-distance. Bootstrap values (1000 replications) are shown at the branches. '**' indicate the sequences generated in this study. Note: The GenBank sequences used in this study represent the introns present in the species and not the species itself.

The second clade included several long introns, grouping in their turn into two clades. The introns of *Bacteriastrum* species, *C. decipiens*, *C. rotosporus* and *C.* cf. *pseudocrinitus* were recovered in the first of these (62%), all of these introns were present at position three in the 18S alignment (Fig. 6.1.2). This clade contained also a group 1 intron, subgroup IC1 in the 18S rDNA of *Aureoumbra lagunensis* (GenBank No. U40258). The sister clade included introns of *Bathycoccus* species, and the basal structure of these clades was poorly resolved.

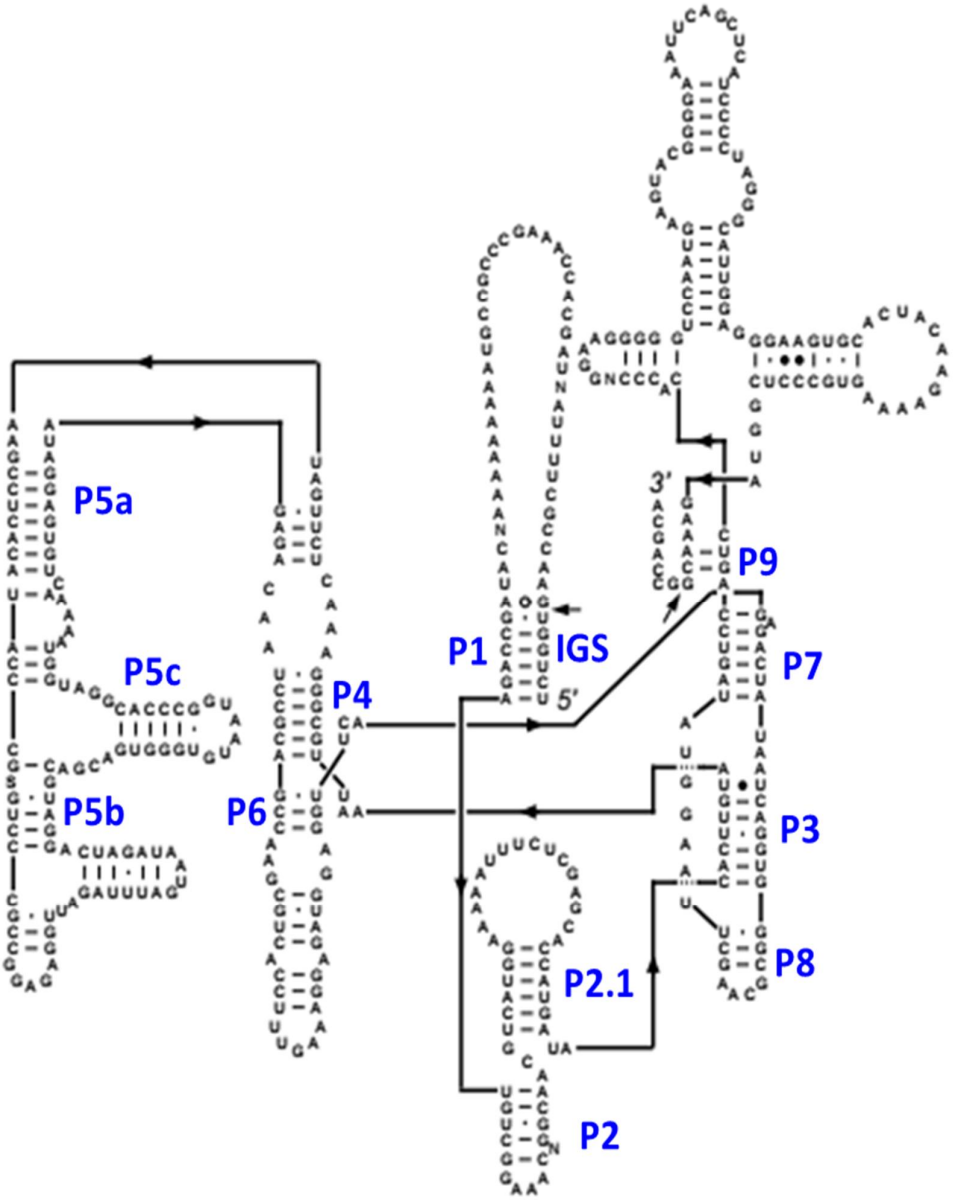


Fig. 6.4.1. Secondary structure model of group IC1 intron in the 18S rRNA of *Aureoumbra lagunensis* (GenBank No. U40258) reproduced from Bhattacharya *et al.* (2001). The arrows mark the 5' and 3' splice junctions; so are the different paired segments P1-P9 and the internal guide sequence (IGS). The secondary model was retrieved from the CRW website.

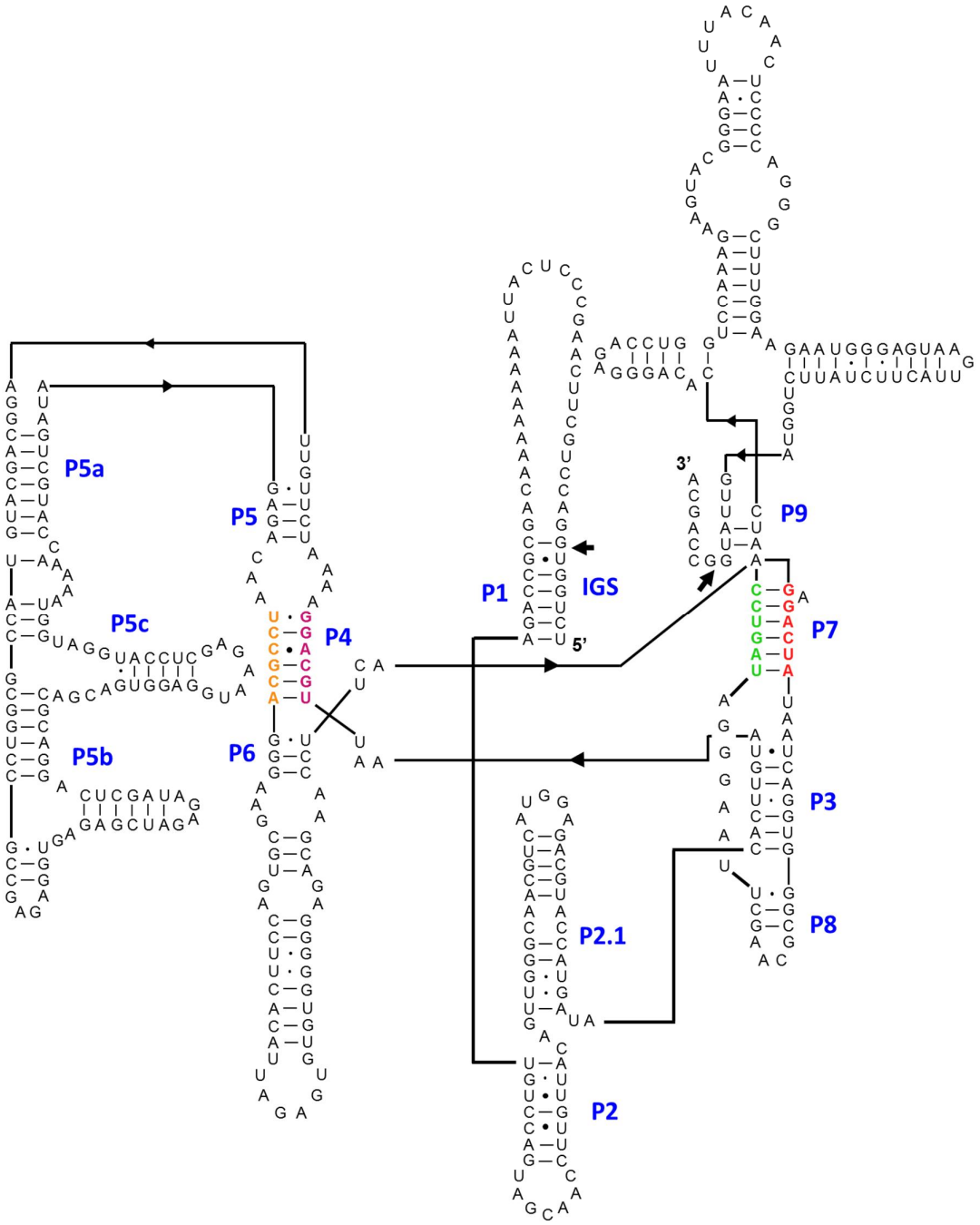


Fig. 6.4.2. Secondary structure of group IC1 intron in the 18S rRNA of *Bacteriastrum* cf. *furcatum* (strain Na8A3). The arrows mark the 5' and 3' splice junctions; so are the different paired segments P1-P9 and the internal guide sequence (IGS). The nucleotides highlighted in the structure represent the conserved sites for P (light purple), Q (orange), R (red) and S (green).

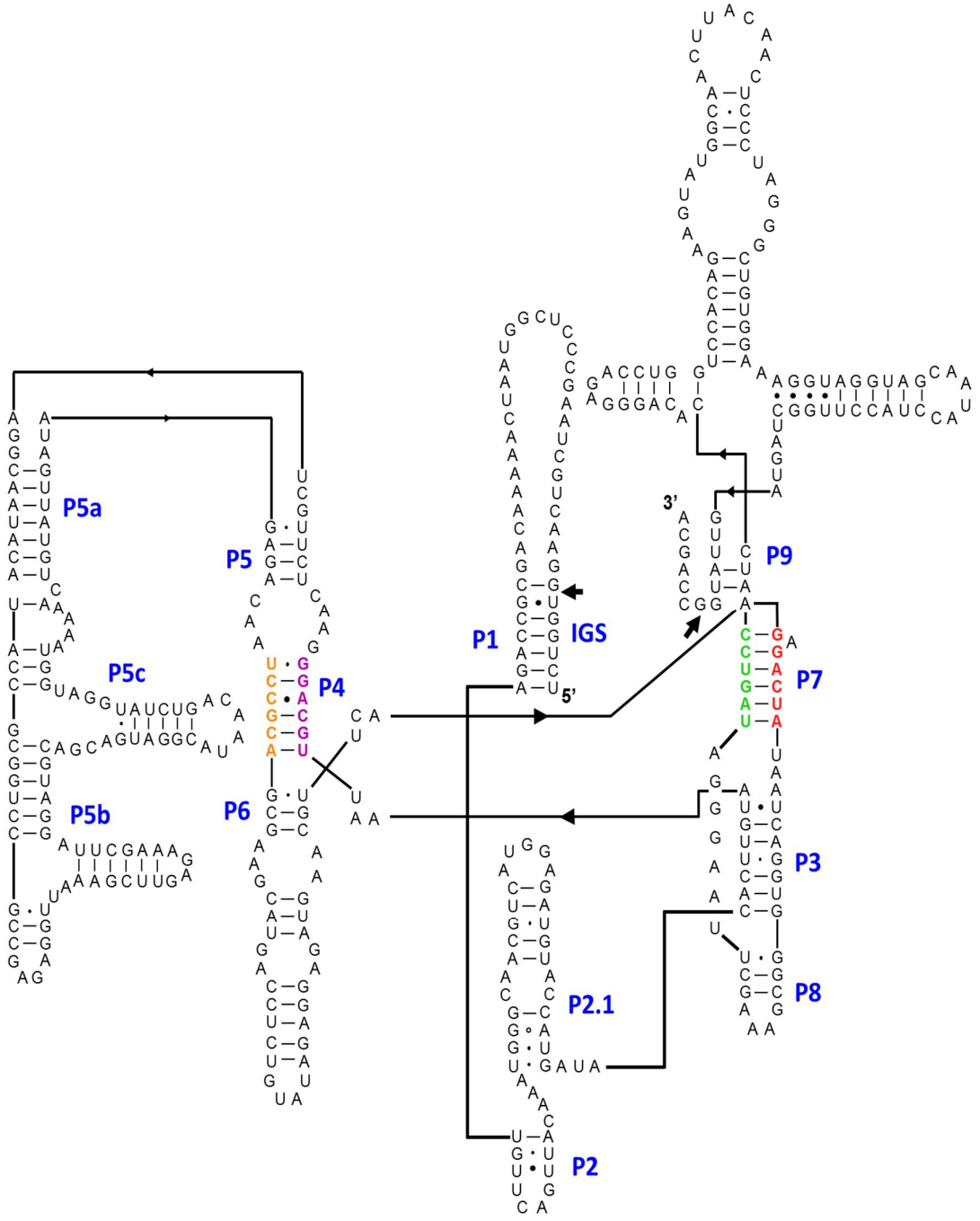


Fig. 6.4.3. Secondary structure of group IC1 intron in the 18S rRNA of *Chaetoceros decipiens* (strain Na1A4). The arrows mark the 5' and 3' splice junctions; so are the different paired segments P1-P9 and the internal guide sequence (IGS). The nucleotides highlighted in the structure represent the conserved sites for P (light purple), Q (orange), R (red) and S (green).

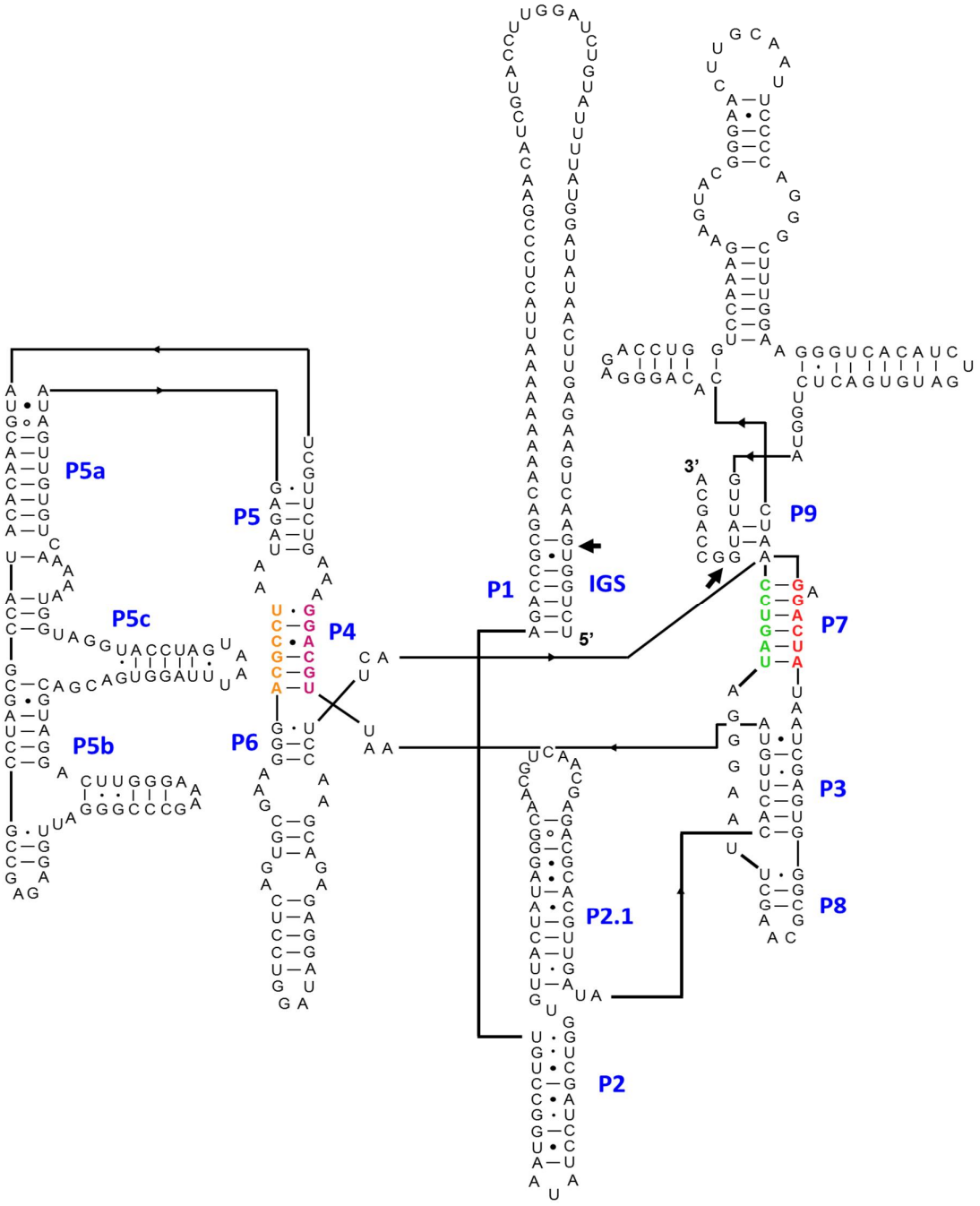


Fig. 6.4.4. Secondary structure of group IC1 intron in the 18S rRNA of *Chaetoceros* cf. *pseudocrinitus* (strain Na12A3). The arrows mark the 5' and 3' splice junctions; so are the different paired segments P1-P9 and the internal guide sequence (IGS). The nucleotides highlighted in the structure represent the conserved sites for P (light purple), Q (orange), R (red) and S (green).

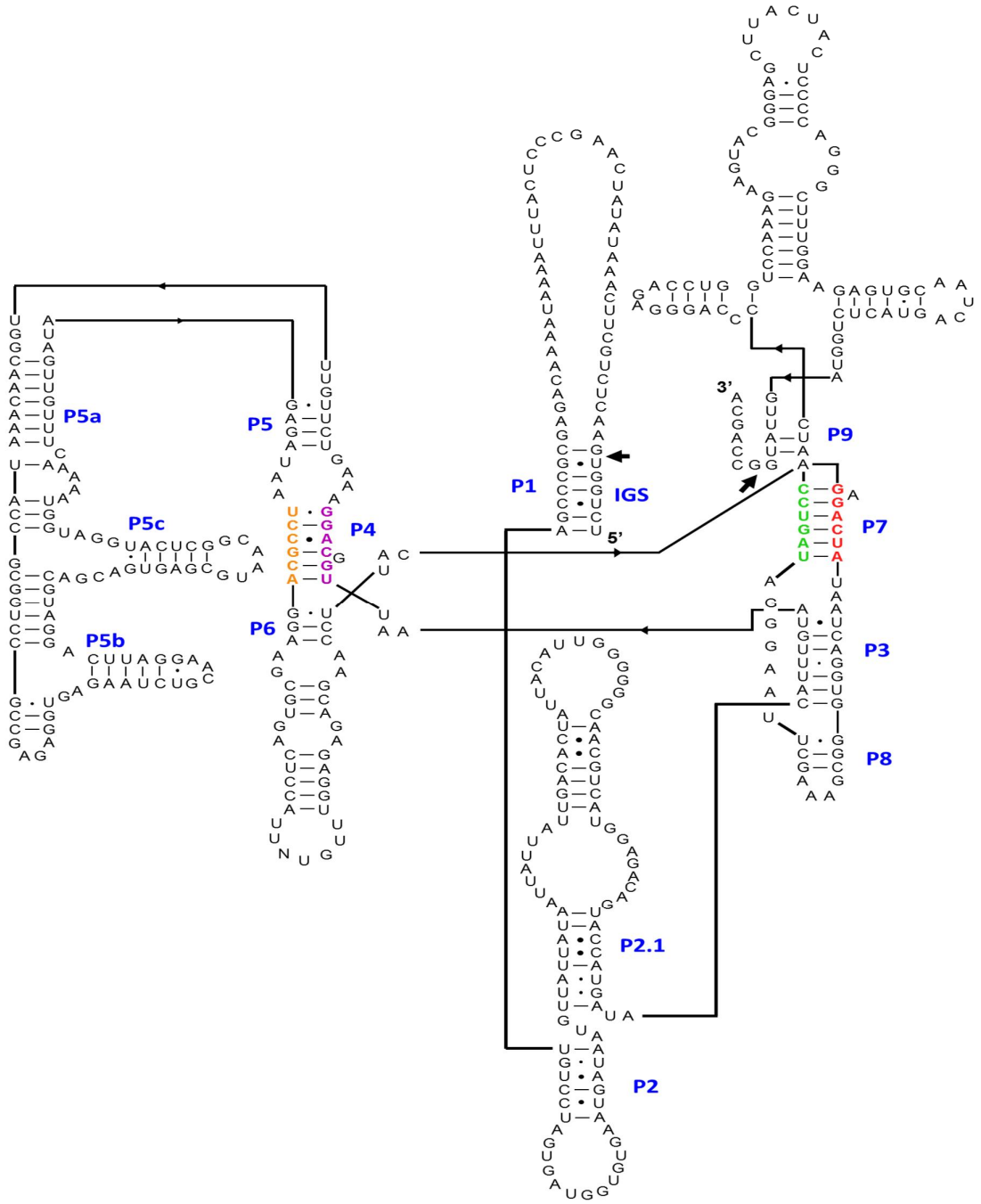


Fig. 6.4.5. Secondary structure of group IC1 intron in the 18S rRNA of *Chaetoceros rotosporus* (strain Na22B3). The arrows mark the 5' and 3' splice junctions; so are the different paired segments P1-P9 and the internal guide sequence (IGS). The nucleotides highlighted in the structure represent the conserved sites for P (light purple), Q (orange), R (red) and S (green).

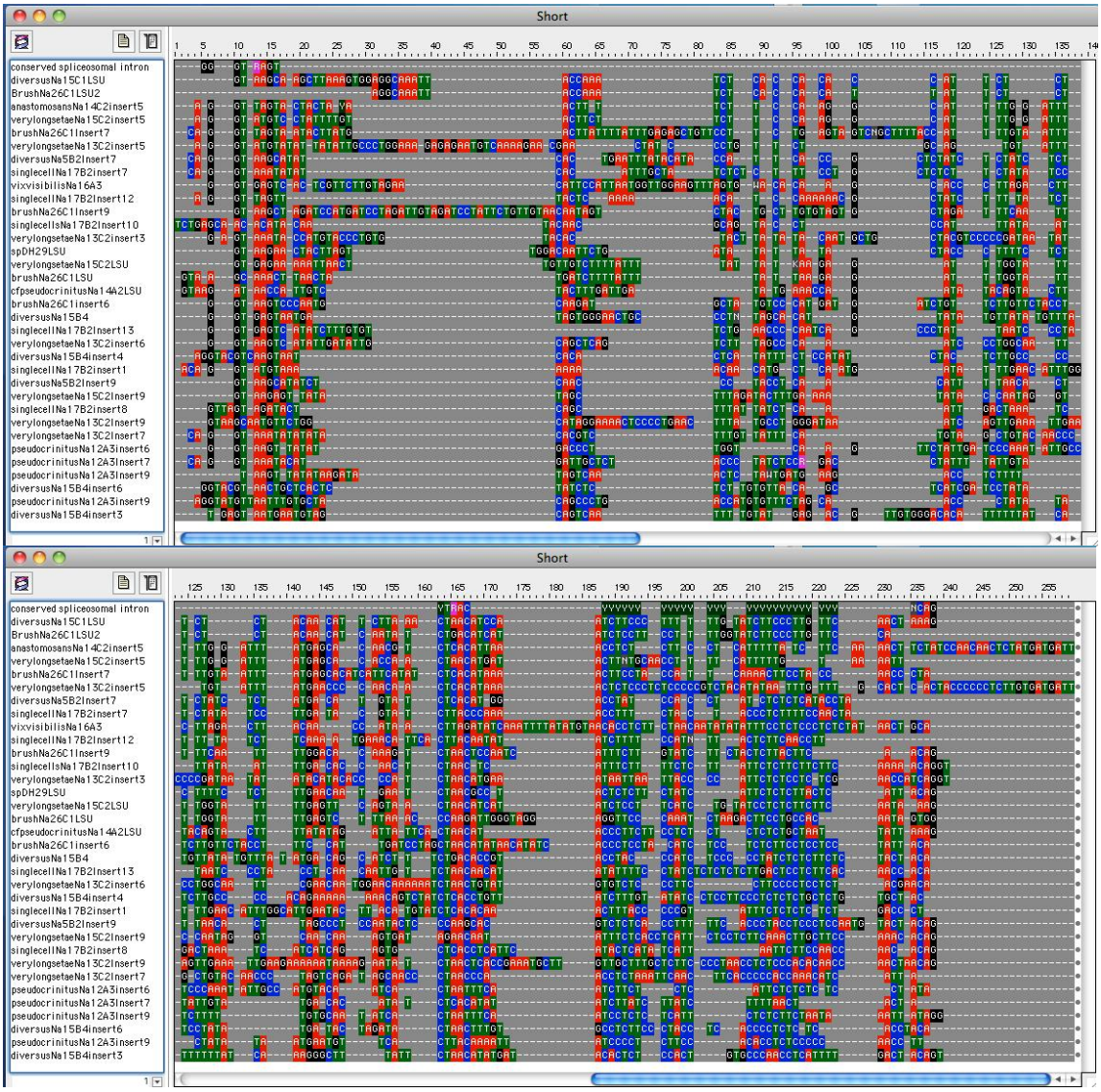


Fig. 6.5. Alignment of the short introns present in the Chaetocerotacean nuclear encoded ribosomal DNA (18S and partial 28S). The first row of the alignment shows the conserved position of the spliceosomal intron. The CT-rich region towards the 3'-end is indicated in blue (C) and green (T).

Primary and secondary structure of the long introns: the secondary structure of the long introns of *Bacteriastrum* species, *C. decipiens*, *C. rotosporus* and *C. pseudocrinitus* present in the second clade were similar to the secondary structure of the group 1 intron, subgroup IC1, in the 18S rDNA sequence of *A. lagunensis.* The secondary structure of the intron of *A. lagunensis* has been depicted in Fig. 6.3.1; that of *Bacteriastrum* cf. *furcatum*, *C. decipiens, C.* cf. *pseudocrinitus* and *C. rotosporus* have been illustrated in Figs 6.3.2–6.2.5.

The P, Q, R, and S sites were located in the obtained model of *A. lagunensis* and later were mapped on the introns in Figs 6.3.2–6.3.5. The result of the plotting exercise of the

Chaetocerotacean intron sequences over the secondary structure of the *A. lagunensis* intron revealed highly similar secondary structure with only some length variation in loops.

Subsequently, the obtained structure of the Chaetocerotacean group I intron of interest was used to write a bpseq file in order to obtain the secondary structure after calculating the thermodynamic entropies. The structure contained three domains as in group I introns. The P1-P2-P10 formed one domain, the second domain was formed of P4-P5-P6 and third domain comprised of P3-P7-P8-P9, as present in all group I introns.

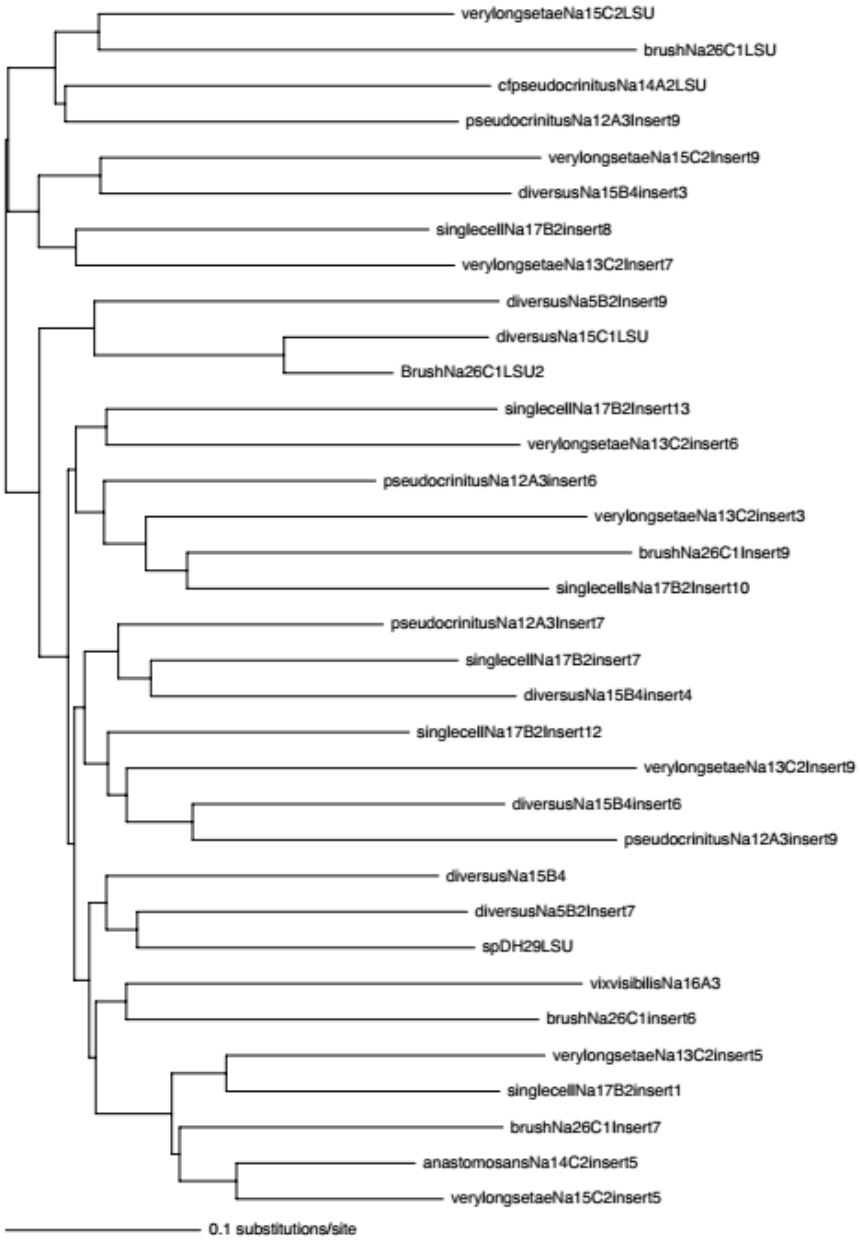


Fig. 6.6. Phylogeny of short introns interrupting the nuclear 18S and 28S rDNA of Chaetocerotacean species. This tree was built with neighbour joining, using pairwise dissimilarity.

Sequence analysis and phylogeny of the short introns: a BLAST search in the GenBank nucleotide collection, using the short introns of Chaetocerotacean strains as a query, did not return any hits. The short intron revealed a highly variable CT-rich site (see Fig. 6.5). The topology of the phylogeny inferred from the short Chaetocerotacean introns (Fig. 6.6) lacked any internal structure, and the terminal nodes were recovered on long branches.

Primary and secondary structure of the short introns: alignment of the short introns showed the following features: a consensus sequence GURAGU at the 5'-end, a highly variable region of ca. 50 bp, a consensus sequence YURAC, a highly variable CT-rich region of ca. 30 bp, and then a consensus sequence ACAG at the 3'-end. Tests with secondary structure folding programs did not return any conserved secondary structures in these short introns; in fact, they returned little or no secondary structure at all.

Discussion

Introns are widespread in the tree of life (Lambowitz & Belfort 1993; Saldanha *et al.* 1993; Yamada *et al.* 1994; Bhattacharya *et al.* 1996a; Müller *et al.* 2001). The results of the present study demonstrate that these introns are also common in particular lineages of Chaetocerotaceae. In fact, there exist two types of introns in the 18S and partial 28S rDNA of the family: group I introns and spliceosomal introns.

The group I introns: the long introns, which vary in length between 360 and 505 bp, exhibit all the sequence characteristics of group 1 introns and their rRNA transcripts fold accordingly. The secondary structure of the group I introns in the family Chaetocerotaceae agrees with the existing models for the secondary structure of group I introns. The conserved P, Q, R, and S regions that form the catalytic core of group I intron (Cech 1988; Michel and Westhof 1990) are also observed in the investigated introns from Chaetocerotacean strains. The GC pair in the stem P7 involved in the guanosine-binding site is considered to be highly conserved (Michel *et al.* 1989) in these introns, whereas in the case of Chaetocerotaceae it was GG pair. And the internal guide sequence (IGS) at the 5' end in my group I introns (i.e. P1 region) are also present. This feature strongly suggests that the long introns found in the Chaetocerotaceae family constitute group I introns. The group 1 introns of the strains *Bacteriastrum* cf. *furcatum* Na8A3, *Chaetoceros decipiens* Na1A4, *C.* cf. *pseudocrinitus* Na12A3 and *C. rotosporus* Na22B3 can even be classified to the subgroup IC1, based on the sequence information and secondary structure as predicted by Bhattacharya *et al.* (2001) because they contain all the characteristics of this group.

According to Bhattacharya *et al.* (1996a), the insertion positions of group 1 introns in the 18S rRNA gene are conserved. With the availability of more data, however, novel positions for introns have been revealed (Gargas *et al.* 1995; Müller *et al.* 2001). The position of the subgroup IC1 introns in Chaetocerotaceae also conflicts with its position in other groups of species for this intron, i.e. at position 516 in the 18S rDNA of *Acanthamoeba, Aureoumbra* and Bangiales species (Bhattacharya *et al.* 2001).

The spliceosomal introns: the short introns are by far the most common ones in the Chaetocerotacean rDNA sequences, and are between 90 and 162 bp in length. They can even occur within other spliceosomal introns (in *C. vixvisibilis* and in *C.* 'singlecells'). These introns are probably spliceosomal introns because they possess all the expected sequence features of this type of intron, namely the GU at the 5' end of the intron, the AG at the 3' end, the CT-rich region upstream from the AG, still further upstream the A embedded in YURAC involved in the bridge formation (Black 2003; Clancy 2008).

The distribution of introns in the rDNA of Chaetocerotaceae: the introns, although common, are not distributed haphazardly all over the Chaetocerotaceae. Instead, they are observed in a total of 14 species in two of the Chaetocerotacean clades. The putative spliceosome introns occur in the *Bacteriastrum* clade in *B. hyalinum*, and in a large clade in *C. anastomosans*, *C.* 'brush', *C. circinalis*, *C. decipiens*, *C. diversus*, *C.* cf. *pseudocrinitus*, *C.* 'singlecells' (Na17B2), *C.* 'verylongsetae' (Na15C2) and *C. vixvisibilis*, whereas the group 1 introns are present in *B*. cf. *furcatum*, *B. jadranum*, *C. brush*, *C. decipiens*, *C.* cf. *pseudocrinitus*, *C. rotosporus*, *C.* 'singlecells' (Na17B2) and in *C. seiracanthus*. Multiple introns can occur at different sites in the rRNA genes of one and the same specimen, as demonstrated in the cases of *C.* 'brush', *C. diversus*, *C.* cf. *pseudocrinitus* and *C.* 'verylongsetae' (Na13C2). This is no exception because the 18S rRNA gene of some lichen-forming fungi can possess multiple small inserts as well (Gargas *et al.* 1995). Several introns are present in the same gene of e.g., *C. diversus*. In *C.* 'brush' and *C.* cf. *pseudocrinitus* different classes of introns occur side by side in the same 18S rDNA. Moreover, even within those clades in which introns are common, not every species has introns and different species exhibit markedly different, apparently unrelated collections of one or more introns, and some species in those clades lack them all together, at least in the rDNA regions I explored. This result fits observations by Mavridou *et al.* (2000) and Creer (2007) in other organismal groups that such introns can be present in some species and absent in others within the same lineage.

The topology of the group I intron phylogeny differs markedly from that of the 18S and partial 28S rDNA phylogenies of Chaetocerotaceae, which is indicative of horizontal transfer. Moreover, spliceosomal introns in different parts of the 18S rDNA core sequences of, for instance, *C. diversus* differ radically, suggesting that these various introns arrived from different sources. Horizontal transfer of the introns (Vader *et al.* 1994; Yamada *et al.* 1994; Belfort & Perlman 1995) seems thus to hold true in Chaetocerotaceae as well.

The observations that the phylogeny of the spliceosomal introns exhibits very little resolution if any at all, that most of the base changes in the tree are recovered along the terminal branches, and that these introns already show variation (expressed in sequence reading difficulties) among different strains sampled from one and the same species, suggests that the functionally unconstrained parts of these introns evolve rapidly. So rapidly that these introns may serve as population genetic markers. For instance, Neapolitan strains of the same species collected in different seasons at times exhibit changes in their bases or in the number of repeats in small microsatellites within the introns. As an example, the presence of inserts in the 18S rDNA of *C. vixvisibilis* strains collected beginning October 2014 but not in those collected beginning July of that year suggests that these strains originate from different populations.

Why some lineages in this family exhibit introns all over the place in their rDNA sequences, whilst other lineages seem devoid of such introns remains to be clarified. Notably, all our strains that exhibit introns in their 18S and or partial 28S rDNA have been collected in the Gulf of Naples; neither the Chilean strains nor the ones from Brittany exhibit introns in their rDNA sequences. The most likely reason is that the occurrence of introns is species- or clade-specific and that the species in which we encountered introns have not been sampled in those other locations. Gathering strains of these species from other locations and screening these for the occurrence of introns would clarify if introns are indeed lineage-specific.

An alternative, though not mutually exclusive, the hypothesis is that introns are more typical of warm water species. Most of the introns found in this study are present in strains collected during the Neapolitan summer and autumn when the water temperature is high. This hypothesis can be tested by screening the rDNA sequences of the tropical as well as cold-temperate Chaetocerotacean diversity, and assessing if the occurrence of introns is correlated with clade or with environmental conditions.

Conclusion and outlooks

The present study revealed the occurrence of a large number of introns in the ribosomal RNA in two of the clades of the marine planktonic diatom family Chaetocerotaceae. As the ribosomal RNA database is continuously updated, more reports of the ribosomal group I introns are being made available. Several of these introns are well characterized based on their secondary structures. This accumulation of the intron data may help to elucidate not only the evolutionary history of these introns but also allow the generation of hypotheses about possible benefits for their hosts. The presence of all these introns in the rDNA comes at a cost because of the extra resources needed to produce these DNA sequences and their RNA products and because of the need for spliceosomes and other tools to remove them from the transcripts. Yet, there are apparently no benefits. It would be interesting to characterize the remaining intron and to investigate if these introns have any evolutionary benefits in the survival strategies of the species.

General Conclusion

The present study focuses on the marine planktonic diatom family Chaetocerotaceae to explore its species diversity, specifically aimed at addressing the following questions:

1. **How well do we know the diversity of Chaetocerotaceae in our study area, the Gulf of Naples?**
2. **How common is cryptic diversity in this family?**
3. **Can spore morphology reveal species diversity? Is it important to know it?**
4. **What is the relationship between *Bacteriastrum* and *Chaetoceros*?**
5. **How important is the need of validated taxonomic reference barcodes for the study of the biodiversity in the family using a metabarcoding approach?**
6. **Can the metabarcoding approach reveal the entire Chaetocerotacean diversity in the Gulf of Naples?**

To answer these questions, a combined culture-based, morphological, and DNA sequencing approache was used to explore species diversity. One of the results constituted a dataset of taxonomically validated Chaetocerotacean reference barcodes. These barcodes were used to "fish" Chaetocerotacean metabarcodes from a huge collection of Eukaryotic ones generated from a series of plankton samples gathered at the Long Term Ecological Research station MareChiara in the Gulf of Naples (GoN) over three consecutive years. Chaetocerotacean metabarcodes and references were then aligned and analysed phylogenetically. In the phylogenetic tree of the reference barcodes and metabarcodes, the references allowed to "taxonomically identify" the various terminal clades, and the accompanying metadata of the samples permitted me to explore the seasonality of these terminal clades. The results of both the combined culture-based, morphological, and DNA sequencing approach and the metabarcoding approach revealed a high diversity of Chaetocerotacean species in the Gulf of Naples with, as expected, highly diverse *Chaetoceros* and less diverse *Bacteriastrum*, as also observed in the morphological observations (Zingone *et al.* 2010; Montresor *et al.* 2013) and in observations elsewhere (Malviya *et al.* 2016).

How many species are characterized in my study? How many of these were already previously described and how many are possibly new to science?

In the present study, 270 strains were isolated from the coasts of Italy (GoN), Chile (Concepción and Las Cruces) and France (Roscoff). These strains were characterised morphologically and genetically into 60 of these species. Sequence information was not obtained from cultured strains of several morphologically defined species due to failure of

amplification. Specimens of three additional species (*C. coarctatus, C. dadayi* and *C. tetrasticon)* were isolated but failed to grow them in culture. The morphology of 50 morphologically distinct species is described in the present study. The remaining species were all cryptic, that is, within the limits of my investigation because of lack of time to investigate several of these "cryptic species" in EM (see Chapter 3).

Among the 60 genetically distinct species investigated in this study, 20 were already characterised morphologically and genetically by other authors (Kooistra *et al.* 2010; Li *et al.* 2013; Bosak *et al.* 2015; Chanmansinp *et al.* 2015). Of the remainder, 15 species were already described morphologically but failed to add genetic information to these species. This means that the morphology of these strains matches that of the material described in the literature. In addition, morphological and genetic information was generated for another 25 species. These were either cryptic, morphologically well-defined and different from anything reported in the literature until the present, or of taxonomic uncertain status. With the latter, I mean that these species were described often more than a century ago and their illustrations are based on light microscopic observation and often sketchy. Strains in this latter category are indicated with "cf." In such cases, these strains need to be compared with the type material in order to resolve the taxonomic uncertainty. The morphological description of 44 species (both confirmed names and "cf.") is provided in Chapter 3. Also, two new species of *C. sporotruncatus* and *C. dichatoensis* are described in Chapter 5.

Are there discrepancies between my observations of strains and species descriptions in the literature?

Not all the strains identified in the present study fit perfectly with descriptions of known species in the literature. Hernández-Becerril *et al.* (2010) amended the description of *C. vixvisibilis* based on field material from the Adriatic Sea. My morphological description of cultured strains of *C. vixvisibilis* from the Gulf of Naples deviates from that of the amended description. Hernández-Becerril *et al.* (2010) mention the presence of a single chloroplast per cell and smooth setae, whereas GoN strains revealed two chloroplasts per cell and setae ornamented with spines. Either the Neapolitan material constitutes a new species or a morphologically distinct population of this species, or cell lines change their morphology in culture. To resolve this issue, one needs to get strains from the Adriatic Sea, characterize these genetically and check their morphology in culture. In addition, one needs to check the morphology of the field material.

In another case, *B.* cf. *furcatum* from the Gulf of Naples and that from the Adriatic Sea showed marked differences in their genetic makeup, resulting in uncertainty as to which one, if any of these at all, is the true *B. furcatum*. The sequences of these two entities are recovered in different parts of the phylogeny, suggesting the convergent evolution of morphological characters defining the species.

How precise species be described using morphology? Is identification always accurate for these species?

In the present study, strains were screened first in LM to assign a provisional name and then characterized the strains genetically. If sets of strains were found to be morphologically highly similar but genetically distinct, ultrastructural details of frustule elements were investigated (and if available, their spore morphology). Hence, molecular and morphological information was integrated to confirm crypticity. This way, potential cryptic diversity was uncovered within several morphologically defined species: *C. affinis, C. brevis, C. contortus*, *C. convolutes, C. curvisetus*, *C. diadema*, *C. debilis*, *C. didymus*, *C. diversus*, *C. lorenzianus*, *C. peruvianus* and *C. socialis*. To define a '*species'*, morphological and molecular characterisation was combined. Groups of strains that were genetic different from other such groups, were considered different species. Phenotypic plasticity among different vegetative life stages can challenge morphological identification; strains may appear morphologically distinct, but the difference may be due merely to different life stages. This was sorted out by assessing their morphological changes over several months in culture and by characterizing them genetically.

Can resting spore morphology be used in species identification?

At times when the vegetative morphology is identical between two groups of genetically different strains, resting spore morphology could be a discriminating character; and this was used to investigate this in the *C. socialis* complex. In fact, in this morphologically defined species, the only differentiating morphological character between four such groups of genetically different strains was the resting spore morphology (Chapter 5). Hence, resting spores can be important in discriminating among species whose vegetative morphology is cryptic.

In this study, resting spore formation was induced in several strains of other species in *Chaetoceros*, especially morphologically defined species revealing cryptic diversity. At times, spores were obtained in culture strains, but this was not under experimental conditions

specifically targeted to obtain spores. Several media described in the literature (Kuwata *et al.* 1993; Sugie & Kuma 2008) were used, along with different light conditions, but these methods did not provide any resting spores in my study (or they did, but the control did as well, or the control did, but the condition supposed to stimulate spore formation failed to generate spores). Hence, factors stimulating spore formation need to be explored, and in various Chaetocerotacean species because different species may have different triggers for spore formation.

What is the phyletic status of the genera* Bacteriastrum *and* Chaetoceros*?

From the phylogenies inferred from 18S and partial 28S rDNA using different algorithms, it is evident that the family Chaetocerotaceae is monophyletic and separates into five distinct clades. The five clades show well-resolved internal structure, whereas the basal structure is poorly resolved. Both the nuclear rDNA gene phylogenies resolve a monophyletic *Bacteriastrum* inside paraphyletic *Chaetoceros,* though with <50% bootstrap support and <95% Bayesian posterior probability. Based on this insufficient support, it cannot be excluded that both these genera are sister species. The poor resolution of the basal ramifications could result from rapid diversification right after the origin of the family.

In the 28S rDNA tree, *Eucampia* resolves inside the *Chaetoceros* clade, whereas in the 18S rDNA tree *Eucampia* resolves as a sister group to the Family Chaetocerotaceae. One reason why *Eucampia* resolves inside the Chaetoceros is probably the scarcity of outgroup sequences in the 28S rDNA phylogeny as compared to the 18S rDNA tree.

Is the V4-region a good reference barcode in metabarcode studies of Chaetocerotaceae?

Before assessing the biodiversity of the family using metabarcoding approach, it was needed to ascertain if the V4 region of the 18S is capable of extracting all and only the Chaetocerotacean metabarcodes from a dataset of eukaryotic metabarcodes. Therefore, the phylogeny from the V4 region was inferred, with the inclusion of some distant outgroups. The phylogeny of the V4 region revealed similar clade structure as that obtained from 18S and from the partial 28S phylogenies, with few changes within internal clades. It also supported the monophyly of Chaetocerotaceae, which indicates that the V4 region is capable of retrieving all the different species in the family in one clade.

In addition, the species that can distinguish using 28S or 18S, can be distinguish also with the V4 region alone. Some of the species for which stains were gathered at multiple geographically distant sites, showed minute sequence differences (one to two base pairs) in

the 18S. Such differences were not always located in the V4 region, meaning that their V4 regions were identical. Therefore, the V4 region is probably too conserved to distinguish between geographic strains within the same species (meta-population).

One objective of this study was to resolve the taxonomic uncertainty due to the presence of cryptic species and species new to science. Therefore, several strains were gathered from different seasons and characterised them morphologically and molecularly to give them a name. Marker sequences of these taxonomically validated strains can then be used as reference barcodes in metabarcoding studies. These reference barcodes also help setting similarity threshold cut-off values of pairs of sequences in an HTS analysis, as in most studies to date; such thresholds are set without knowing the minimum genetic difference among biologically and/or morphologically distinct species. In this study, the reference barcode region of V4 was able to differentiate species at the 100% threshold. If similarity threshold was decreased, loss of species diversity was observed.

How many species were detected in the HTS approach?

An objective of the study was to assess whether metabarcoding approaches can accurately monitor Chaetocerotacean diversity down to the species level. The metabarcoding approach with high-throughput sequencing (HTS) further enhanced our knowledge of the species diversity in the GoN over the seasonal cycle; 46 species were identified in my culture approach whereas a total of 30 additional species were recovered in the metabarcode analysis, leading to a total of 76 Chaetocerotacean species. Of course, it needs to be taken into account that sequences for ca. 19 Neapolitan species were not obtained due to failures and three species could not be cultured in the first place. Some of the species detected in the HTS data were not known to occur in GoN, such as *C. cinctus*, *C.* cf. *constrictus* and *C. debilis*. Another surprising finding is that some of the unknown species appear apparently in high numbers. Hence, it will be of interest to see what these hidden species are.

Are the V4 primers able to amplify the V4 regions of all Chaetocerotaceae?

The V4 universal primers were evaluated for the primer misfits, and from this investigation, it was found that the V4 primers are not universal to all the members of the family Chaetocerotaceae. In the case of *C. radicans*, the primer show critical mismatches leading to loss of species diversity in the metabarcoding analysis. However, it cannot be conclude that this hampers the PCR amplifications of this species' V4-region in the metabarcoding approach, as it is difficult to prove that these species are absent in the GoN. Also, mismatches

were seen in a few other species, but in spite of that, some of these species showed up in the metabarcoding data. Hence, to resolve this problem, one needs to validate the primer compatibility by doing trials with *C. radicans* or to study those localities that harbour this species.

Why is it important to study the introns in Chaetocerotaceae?

In the present study, introns were reported in the 18S and 28S rDNA gene regions (Chapter 6). Some 18S rDNA sequences contained multiple introns; in that of *C. diversus*, there were eight. The partial 28S alignment showed only one region with a short intron in *C. diversus* and one in *C. '*brush'*,* whereas the 18S alignment had twelve such regions showing an intron; three of these regions contained a long intron (ca. 450 bp) and nine contained a short one (ca. 150 bp). Some of these introns were present in the primers sites of amplification or sequencing primers, thus affecting amplification and/or sequencing of these regions. Also in a few cases, these introns were present in the V4-region, i.e., the region used in the metabarcoding analysis. If a species shows an intron in the V4-region, then it will not show up in the HTS data. Hence, it is important to know the compatibility between primers and their intended target regions and be aware of the possible presence of introns in the target regions for diversity assessments when using the metabarcoding approach.

The long introns observed in the present study belong to the subgroup IC1 of the group I introns (Bhattacharya *et al.* 2001). They contain the conserved domains for the P, Q, R, and S sites, which are the marker sites for the group I introns. These long introns are present in two major clades (the *Bacteriastrum* clade and the clade including *C. '*brush'*, C. decipiens, C. diversus, C. rotosporus* and *C. seiracanthus*. The phylogeny of the long introns does not corroborate the phylogeny of Chaetocerotaceae, suggesting that these introns have evolved independently from their host-sequences, i.e., lateral "gene" transfer. The short introns show high variation, both in length and base composition, so that only certain conserved regions can be aligned with some confidence. It is still possible to see that these introns all belong to a "family" because they show regions with similar base composition. Yet, the phylogeny was unresolved with most of the base changes observed in the end branches, suggesting that homoplasy has obliterated most of the phylogenetic information. Support for this hypothesis comes from Sanger sequences obtained from strains belonging to the same species. Some of these sequences showed ambiguities and reading difficulties in the intron region, indicative of intraspecific variation, fast evolving regions.

Future perspectives

Several possibilities for future research are opened by the result of my work.

- A detailed investigation is needed on those taxa that failed to amplify upon retrieval in culture. Alternate methods like, single cell amplification or nested PCR can be used to study this unexplored diversity. The single cell approach will provide genetic information for those species that are uncultivable in laboratory conditions.
- Several of the species occurring sympatrically showed cryptic diversity resulting in the nomenclatural problems. The allocation of the type material and the type specimens needs attention, and this work can be a baseline for such work.
- The resting spores of the members of the family reveal a wide range of ornamentation. How these single-celled species can generate such enormous range of ornamentation is still unaddressed. One example is the resting spore of *C. lorenzianus*, in which there are two prominent conical elevations with dichotomous branching process on the primary valve and two elevations on the secondary valve.
- Morphological and genetic characterisation is needed for species that are not included in my study. The species *C. bacteriastroides* Karsten is of interest as this species might be sister to the *Bacteriastrum*-clade; it shares the presence of multiple chloroplasts per cell and possesses more than two setae per valve, although the extra setae are small, do not fuse, and appear ultrastructurally different from the large setae.
- The reference barcodes generated in the present study can roughly allow microbial ecologist to understand the Chaetocerotacean diversity in their own geographic area. Although, as stated in Chapter 4, the reference barcode may or may not be identical to the dominant haplotypes in different geographic regions, but it will at least reveal the taxonomic identity of the species in their regions.
- The results obtained from the metabarcoding exercise revealed several species that are unknown and need attention. Work was initiated to collect these species based on seasonality and their occurrence in the HTS data.
- The introns obtained in the present study need a proper investigation to understand the importance of these introns in the family Chaetocerotaceae. What is the significance of the introns and why are they only present in a few groups of species?
- A proper study is needed to investigate resting spore formation in the Chaetocerotaceae.
- The functions of the setae need to be investigated.

Bibliography

Aké Castillo J.A., Guerra-Martínez S.L. & Zamudio-Reséndiz M.E. (2004) Observations on some species of *Chaetoceros* (Bacillariophyceae) with reduced number of setae from a tropical coastal lagoon. *Hydrobiologia*, 1: 203−213.

Alberts B. (2008) Molecular biology of the cell. Garland Science. New York.

Alva V., Nam S.Z., Söding J. & Lupas A.N. (2016) The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Research*, 44: 410−415.

Alverson A.J. & Theriot E.C. (2005) Comments on recent progress towards reconstructing the diatom phylogeny. *Journal of Nanoscience & Nanotechnology*, 5: 57−62.

Alverson A.J. (2007) Strong purifying selection in the silicon transporters of marine & freshwater diatoms. *Limnology & Oceanography*, 52: 1420−1429.

Alverson A.J., Jansen R.K., & Theriot E.C. (2007) Bridging the Rubicon: Phylogenetic analysis reveals repeated colonizations of marine & fresh waters by Thalassiosiroid diatoms. *Molecular Phylogenetics and Evolution*, 45: 193−210.

Alverson A.J. (2008) Molecular systematics and the diatom species. *Protist*, 159: 339−353.

Amaral-Zettler L.A., McCliment E.A., Ducklow H.W. & Huse S.M. (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small−subunit ribosomal RNA genes. *PLoS One*, 4(7): e6372.

Amato A., Kooistra W.H., Ghiron J.H., Mann D.G., Pröschold T. & Montresor M. (2007) Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist*, 158: 193−207.

Amato A. & Montresor M. (2008) Morphology, phylogeny, and sexual cycle of *Pseudo-nitzschia mannii* sp. nov. (Bacillariophyceae): a pseudo-cryptic species within the *P. pseudodelicatissima* complex. *Phycologia*, 47: 487−497.

Anderson N.J. & Battarbee R.W. (1994) Aquatic community persistence and variability: a palaeolimnological perspective. Blackwell Scientific Publications, Oxford.

Andersson A.F., Riemann L. & Bertilsson S. (2010) Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *The ISME Journa*l, 4: 171–181.

Antczak M., Zok T., Popenda M., Lukasiak P., Adamiak R.W., Blazewicz J. & Szachniuk M. (2014) RNApdbee - a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs. *Nucleic Acids Research*, 42: W368–W372.

Appeltans W., Ahyong S.T., Anderson G., Angel M.V., Artois T., Bailly N., Bamber R., Barber A., Bartsch I., Berta A. & Błażewicz-Paszkowycz M., *et al.* (2012) The magnitude of global marine species diversity. *Current Biology*, 22: 2189–2202.

Ashworth M.P., Nakov T. & Theriot E.C. (2013) Revisiting Ross & Sims (1971): toward a molecular phylogeny of the Biddulphiaceae and Eupodiscaceae (Bacillariophyceae). *Journal of Phycology*, 49: 1207–1222.

Assmy P., Hernández-Becerril D.U. & Montresor M. (2008) Morphological variability and life cycle traits of the type species of the diatom genus *Chaetoceros*, *C. dichaeta. Journal of Phycology*, 44: 152–163.

Assmy P. & Smetacek V. (2012) Algal blooms: In Topics in Ecological & Environmental Microbiology. Elsevier, San Diego, pp. 85−100.

Avaria S., Palma S., Sievers H. & Silva N. (1989) Revisión sobre aspectos oceanográficos físicos, químicos y planctológicos de la bahía de Valparaso y áreas adyacentes. *Biología Pesquera*, 18: 67−96.

Balzano S, Sarno D. & Kooistra W. (2011) Effects of salinity on the growth rate and morphology of ten *Skeletonema* strains. *Journal of Plankton Research*, 33: 937−945.

Behnke A., Friedl T., Chepurnov V.A. & Mann D.G. (2004) Reproductive compatibility and rDNA sequence analyses in the *Sellaphora pupula* species complex (Bacillariophyta) 1. *Journal of Phycology*, 40: 193−208.

Behnke A., Barger K.J., Bunge J. & Stoeck T. (2010) Spatio-temporal variations in protistan communities along an O_2/H_2S gradient in the anoxic Framvaren Fjord (Norway). *FEMS Microbiology Ecology*, 72: 89−102.

Behrenfeld M.J. (2010) Abandoning Sverdrup's Critical Depth Hypothesis on phytoplankton blooms. *Ecology*, 91: 977−989.

Belfort M. & Perlman S. (1995) Mechanisms of intron mobility. *Journal of Biological Chemistry*, 270: 30237–30240.

Belfort M. & Roberts R.J. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Research*, 25: 3379–3388.

Berg J.M., Tymoczko J.L. & Lubert S. (2007) Biochemistry (6 ed.), New York, USA

Berger S.A., Krompass D. & Stamatakis A. (2011) Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Systematic Biology*, 60: 291–302.

Berget S.M., Moore C. & Sharp P.A. (1977) Spliced segments at the 5’ terminus of adenovirus 2 late mRNA. *Proceedings of National Academy of Sciences of the USA*, 74: 3171−3175.

Bergkvist J., Thor P., Jakobsen H. & Selander E. (2012) Grazer-induced chain length plasticity reduces grazing risk in a marine diatom. *Limnology and Oceanography*, 57: 318–324.

Beszteri B., John U. & Medlin L.K. (2007) An assessment of cryptic genetic diversity within the *Cyclotella meneghiniana* species complex (Bacillariophyta) based on nuclear and plastid genes, and amplified fragment length polymorphisms. *European Journal of Phycology*, 42: 47–60.

Bhattacharya D., Friedl T. & Damberger S. (1996a) Nuclear encoded rDNA group-I introns; origin and phylogenetic relationships of insertion site lineages in the green algae. *Molecular Biology and Evolution*, 13: 978–989.

Bhattacharya D., Damberger S., Surek B. & Melkonian M. (1996b) Primary and secondary structure analyses of the rDNA group-I introns of the Zygnematales (Charophyta). *Current Genetics*, 29: 282–286.

Bhattacharya D., Lutzoni F., Reeb V., Simon D., Nason J. & Fernandez F. (2000) Widespread occurrence of spliceosomal introns in the rDNA genes of ascomycetes. *Molecular Biology and Evolution*, 17: 1971–1984.

Bhattacharya D., Cannone J.J. & Gutell R.R. (2001) Group I intron lateral transfer between red and brown algal ribosomal RNA. *Current Genetics*, 40: 82–90.

Biggs B.J.F. & Hickey C.W. (1994) Periphyton responses to a hydraulic gradient in a regulated river in New Zealand. *Freshwater Biology*, 32: 49–59.

Bittner L., Gobet A., Audic S., Romac S., Egge E., Santini S., Ogata H., Probert I., Edvardsen B. & de Vargas C. (2013) Diversity patterns of uncultured haptophytes unravelled by pyrosequencing in Naples Bay. *Molecular Ecology*, 22: 87–101.

Black D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry*, 72: 291–336.

Booth B.C., Larouche P., Bélanger S., Klein B., Amiel D. & Mei, Z.P. (2002) Dynamics of *Chaetocerossocialis* blooms in the North Water. *Deep Sea Research Part* II. *Topical Studies in Oceanography*, 49: 5003−5025.

Borchardt M.A. (1996) Algal Ecology: Freshwater Benthic Ecosystems Academic Press Inc., California, pp. 184−228.

Bosak S., Pletikapic G., Hozic A., Svetlicic V., Sarno D. & Vilicic D. (2012) A novel type of colony formation in marine planktonic diatoms revealed by atomic force microscopy. *PLoS One*, 7: e44851.

Bosak S., Supraha L., Nanjappa D., Kooistra W.H.C.F. & Sarno D. (2015) Morphology and phylogeny of four species from the genus *Bacteriastrum* (Bacillariophyta). *Phycologia*, 54: 130–148.

Bougoure D.S. & Cairney J.W. (2005) Fungi associated with hair roots of *Rhododendronlochiae* (Ericaceae) in an Australian tropical cloud forest revealed by culturing and culture-independent molecular methods. *Environmental Microbiology*, 7: 1743−1754.

Boyer S., Wratten S.D., Holyoake A., Abdelkrim J. & Cruickshank R.H. (2013) Using Next. Generation Sequencing to analyze the diet of a highly endangered land snail (*Powelliphantaaugusta*) feeding on endemic earthworms. *PLoS One*, 8: e75962.

Buschmann A.H., Riquelme V.A., Hernández-González M.C., Varela D., Jiménez J.E., Henríquez L.A., Vergara P.A., Guíñez R. & Filún L., (2006) A review of the impacts of salmonid farming on marine coastal ecosystems in the southeast Pacific. *ICES Journal of Marine Science*: *Journal du Conseil*, 63: 1338–1345.

Calbet A. & Landry M. (2004) Phytoplankton growth, microzooplankton grazing, & carbon cycling in marine systems. *Limnology & Oceanography*, 49: 51−57.

Cannone J.J., Subramanian S., Schnare M.N., Collett J.R., D'Souza L.M., Du Y., Feng B., Lin N., Madabusi L.V., Müller K.M., Pande N., Shang Z., Yu N. & Gutell R.R. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, 3: 2.

Casteleyn G., Chepurnov V.A., Leliaert F., Mann D.G., Bates S.S., Lundholm N., Rhodes L., Sabbe K. & Vyverman W. (2008) *Pseudo-nitzschia pungens* (Bacillariophyceae): A cosmopolitan diatom species? *Harmful Algae*, 7: 241–257.

Cech T.R. (1986) A model for the RNA-catalyzed replication of RNA. *Proceedings of the National Academy of Sciences USA*, 83: 4360–4363.

Cech T.R. (1988) Conserved sequences and structures of group I introns: building an active site for RNA catalysis-a review. *Gene*, 73: 259–271.

Cech T.R, Damberger S.H. & Gutell R.R. (1994) Representation of the secondary and tertiary structure of group I introns. *Nature Structural Biology*, 1: 273–280.

Chamnansinp A., Li Y., Lundholm N. & Moestrup Ø. (2013) Global diversity of two widespread, colony-forming diatoms of the marine plankton, *Chaetoceros socialis* (syn. *C. radians*) & *Chaetoceros gelidus* sp. nov. *Journal of Phycology*, 49: 1128−1141.

Chamnansinp A., Moestrup Ø. & Lundholm N. (2015) Diversity of the marine diatom *Chaetoceros* (Bacillariophyceae) in Thai waters-revisiting *Chaetoceros compressus* and *Chaetoceros contortus*. *Phycologia*, 54: 161−175.

Chanfreau G. & Jacquier A. (1993) Interaction of intronic boundaries is required for the second splicing step efficiency of a group II intron. *EMBO Journal*, 12: 5173–5180.

Chanfreau G. & Jacquier A. (1996) An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO Journal*, 15: 3466–3476.

Chiswell S.M. (2011) The spring phytoplankton bloom: don't abandon Sverdrup completely. *Marine Ecology Progress Series*, 443: 39−50.

Choi H.G., Joo H.M., Jung W., Hong S.S., Kang J.S. & Kang S.H. (2008) Morphology and phylogenetic relationships of some psychrophilic polar diatoms (Bacillariophyta). *Nova Hedwig Beih*, 133: 7−30.

Chow L.T., Gelinas R.E., Broker T.R. & Roberts R.J. (1977) An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell*, 12: 1–8.

Churro C.I., Carreira C.C., Rodrigues F.J., Craveiro S.C., Calado A.J., Casteleyn G. & Lundholm N. (2009) Diversity and abundance of potentially toxic *Pseudo-nitzschia* Peragallo in Aveiro coastal lagoon, Portugal and description of a new variety, *P. pungens* var. *aveirensis* var. nov. *Diatom Research*, 24: 35−62.

Clancy S. (2008). RNA Splicing: Introns, Exons and Spliceosome. *Nature Education*, 1: 31.

Clément A. & Lembeye G. (1993) Toxic phytoplankton blooms in the sea. Elsevier, Amsterdam, pp. 223−228.

Cleve P.T. (1896) Diatoms from Baffins Bay and Davis Strait. *K. Svenska Vet.-Akad. Handlingar*, 22: 3–22.

Cong J., Yang Y., Liu X., Lu H., Liu X., Zhou J., Li D., Yin H., Ding J. & Zhang Y. (2015) Analyses of soil microbial community compositions and functional genes reveal potential consequences of natural forest succession. *Scientific Reports*, 5: 10007.

Copertino D.W. & Hallick R.B. (1993) Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. *Trends in Biochemical Sciences*, 18: 467–471.

Corsby L.H. & Wood E.J.F. (1958) Studies on Australian and New Zealand diatoms I. - Planktonic and allied species. *Transactions of the Royal Society of New Zealand*, 85: 483–530.

Crease T.J. & Taylor D.J. (1998) The origin and evolution of variable-region helices in V4 and V7 of the small-subunit ribosomal RNA of branchiopod crustaceans. *Molecular Biology and Evolution*, 15: 1430–1446.

Creer S. (2007) Choosing and using introns in molecular phylogenetics. *Evolutionary Bioinformatics Online*, 3: 99–108.

Cullings K.W. (1992) Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. *Molecular Ecology*, 1: 233–240.

Cupp E.E. (1943) Marine plankton diatoms of the west coast of North America. Scripps Institution of Oceanography, University of California, San Diego.

Dai L. & Zimmerly S. (2002) Compilation and analysis of group II intron insertions in bacterial genomes: Evidence for retroelement behavior. *Nucleic Acids Research*, 30: 1091–1102.

Damsté J.S.S., Muyzer G., Abbas B., Rampen S.W., Massé G., Allard W.G., Belt S.T., Robert J.M., Rowland S.J., Moldowan J.M. & Barbanti S.M. (2004) The rise of the rhizosolenid diatoms. *Science*, 304: 584–587.

deVargas C., Audic S., Henry N., Decelle J., Mahé F., Logares R., Lara E., Berney C., Bescot N.L., Probert I., Carmichael M., Poulain J., Romac S., Colin S., Aury J.M., Bittner L., Chaffron S., Dunthorn M., Engelen S., Flegontova O., Guidi L., Horák A., Jaillon O., Lima-Mendez G., Lukeš J., Malviya S., Morard R., Mulot M., Scalco E., Siano R., Vincent F., Zingone A., Dimier C., Picheral M., Searson S., Kandels-Lewis S., Acinas S.G., Bork P., Bowler C., Gorsky G., Grimsley N., Hingamp P., Iudicone D., Not F., Ogata H., Pesant S., Raes J., Sieracki M.E., Speich S., Stemmann L.,

Sunagawa S., Weissenbach J., Wincker P. & Karsenti E. (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348: 6237.

Decelle J., Romac S., Sasaki E., Not F., & Mahé F. (2014) Intracellular diversity of the V4 and V9 regions of the 18S rRNA in marine protists (Radiolarians) assessed by high-throughput sequencing. *PLoS One*, 9: e104297.

Degerlund M. & Eilertsen H.C. (2010) Main species characteristics of phytoplankton spring blooms in NE Atlantic and arctic waters. *Estuarine Coasts*, 33: 242−269.

Degerlund M., Huseby S., Zingone A., Sarno D. & Landfald B. (2012) Functional diversity in cryptic species of *Chaetoceros socialis* Lauder (Bacillariophyceae). *Journal of Plankton Research*, 34: 416−431.

Delegrange A., Vincent D., Courcot L. & Amara R. (2015) Testing the vulnerability of juvenile sea bass (*Dicentrarchuslabrax*) exposed to the harmful algal bloom (HAB) species *Pseudo-nitzschia delicatissima*. *Aquaculture*, 437: 167−174.

Delong E.F. & Pace N.R. (1991) Environmental diversity of bacteria and archaea. *Systematic Biology*, 50: 470–478.

DeNicola D.M. (1996) Algal Ecology: Freshwater Benthic Ecosystems. Academic Press, San Diego, pp. 149–181.

Derot J., Schmitt F.G., Gentilhomme V. & Morin P. (2016) Correlation between long-term marine temperature time series from the eastern and western English Channel: Scaling analysis using empirical mode decomposition. *Comptes Rendus Géoscience*, 348: 343–349.

Dixit S.S., Cumming B.F., Birks H.J.B., Smol J.P., Kingston J.C., Uutala A.J., Charles D.F. & Camburn K.E. (1993) Diatom assemblages from Adirondack lakes (New York, USA) and the development of inference models for retrospective assessment. *Journal of Paleolimnology*, 8: 27–47.

Doudna J.A. & Cech T.R. (2002) The chemical repertoire of natural ribozymes. *Nature*, 418: 222–228.

Doyle J.J. & Doyle J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin*, 19: 11–15.

Dujon B. (1989) Group I introns as mobile genetic elements: facts and mechanistic speculations - a Review. *Gene*, 82: 91–114.

Dujon B., Belfort M., Butow R.A., Jacq C., Lemieux C., Perlman P.S. & Vogt V.M. (1989) Mobile introns: definition of terms and recommended nomenclature. *Gene*, 82: 115–118.

Durbin E.G. (1978) Aspects of the biology of resting spores of *Thalassiosira nordenskioeldii* and *Detonula confervacea*. *Marine Biology*, 45: 31–37.

Edgar R.C., Haas B.J., Clemente J.C., Quince C. & Knight R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27: 2194–2200.

Edgcomb V., Orsi W., Bunge J., Jeon S., Christen R., Leslin C., Holder M., Taylor G.T., Suarez P., Varela R. & Epstein S. (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *The ISME Journal*, 5: 1344−1356.

Edler L. & Elbrächter M. (2010) Microscopic and molecular methods for quantitative phytoplankton analysis. UNESCO Publishing, Paris, pp. 13–20.

Edlund M.B. & Stoermer E.F. (1993) Resting spores of the freshwater diatoms *Acanthoceras* and *Urosolenia*. *Journal of Paleolimnology*, 9: 55–61.

Ehara M., Inagaki Y., Watanabe K.I. & Ohama T. (2000) Phylogenetic analysis of diatom coxI genes and implications of a fluctuating GC content on mitochondrial genetic code evolution. *Current Genetics*, 37: 29–33.

Ehrenberg C.G. (1844) Einige vorläufige Resultate seiner Untersuchungen der ihm von der Südpolreise des Captain Ross, so wie von den Herren Schayer und Darwin zugekommenen Materialien über das Verhalten des kleinsten Lebens in den Oceanen und den grössten bisher zugänglichen Tiefen des Weltmeeres. *Bericht über die zur Bekanntmachung Geeigneten Verhandlungen Der Königl. Preuss. Akademie Der Wissenschaften zu Berlin*, 1: 182−207.

Eiler A., Drakare S., Bertilsson S., Pernthaler J., Peura S., Rofner C., Simek K., Yang Y., Znachor P. & Lindström E.S. (2013) Unveiling distribution patterns of freshwater phytoplankton by a next generation sequencing based approach. *PLoS One*, 8: e53516.

Eilertsen H.C., Sandberg S. & Tøllefsen H. (1995) Photoperiodic control of diatom spore growth; a theory to explain the onset of phytoplankton blooms. *Marine Ecology Progress Series*, 116: 303–307.

Einvik C., Nielsen H., Westhof E., Michel F. & Johansen S. (1998) Group I-like ribozymes with a novel core organization perform obligate sequential hydrolytic cleavages at two processing sites. *RNA*, 4: 530–541.

Eppley R.W. & Harrison W.G. (1975) Proceeding of First International Conference on Toxic dinoflagellate Blooms. Wakefield, pp. 1l−22.

Evans K.M. Wortley A.H. & Mann D.G. (2007) An assessment of potential diatom "barcode" genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). *Protist*, 158: 349–364.

Evans K.M., Wortley A.H., Simpson G.E., Chepurnov V.A. & Mann D.G. (2008) A molecular systematic approach to explore diversity within the *Sellaphorapupula* species complex (Bacillariophyta) 1. *Journal of Phycology*, 44: 215–231.

Evensen D.L. & Hasle G.R. (1975) The morphology of some *Chaetoceros* (Bacillariophyceae) species as seen in the electron microscopes. *Nova HedwigiaBeiheft*, 53: 153−184.

Falkowski P.G., Katz M.E., Knoll A.H., Quigg A., Raven J.A., Schofield O. & Taylor F.J.R. (2004) The evolution of modern eukaryotic phytoplankton. *Science*, 305: 354−360.

Faúndez P.B., Morales C.E., Arcos D. (2001) Variabilidad espacial y temporal en la hidrografía invernal del sistema de bahías frente a la VIII región (Chile centro-sur). *Revista chilena de historia natural*, 74: 817−831.

Ferrario M., Hernández-Becerril D.U. & Garibotti I. (2004) Morphological study of the marine planktonic diatom *Chaetoceroscastracanei* Karsten (Bacillariophyceae) from

Antarctic waters, with a discussion on its possible taxonomic relationships. *Botanica Marina*, 47: 349–355.

Field C.B., Behrenfeld M.J., Randerson J.T. & Falkowski P. (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*, 281: 237–240.

Findlay H.S., Yool A., Nodale M. & Pitchford J.W. (2006) Modelling of autumn plankton bloom dynamics. *Journal of Plankton Research*, 28: 209−220.

Foley S., Bruttin A. & Brüssow H. (2000) Widespread distribution of a group I intron and its three deletion derivatives in the lysin gene of *Streptococcus thermophilus* Bacteriophages. *Journal of Virology*, 74: 611–618.

Fonseca V.G., Carvalho G.R., Nichols B., Quince C., Johnson H.F., Neill S.P., Lambshead J.D., Thomas W.K., Power D.M. & Creer S. (2014) Metagenetic analysis of patterns of distribution and diversity of marine meiobenthic eukaryotes. *Global Ecology and Biogeography*, 23: 1293–1302.

French F.W. & Hargraves P.E. (1980) Physiological characteristics of plankton diatom resting spores. *Marine Biology Letters*, 1: 185–195.

Frey B.F. & Suppmann B. (1995) Demonstration of the Expand PCR System's greater fidelity and higher yields with a lacI–based fidelity assay. *Biochemica*, 2: 34–35.

Fritz S.C., Cummings B.C., Gasse F. & Laird K.F. (1999) The Diatoms: applications for the environmental and earth sciences. Cambridge University Press. Cambridge, pp. 41–73.

Fryxell G.A. (1978) Chain-forming diatoms: three species of Chaetocerotaceae. *Journal of Phycology*, 14: 62−71.

Fryxell G.A. & Hasle G.R. (2003) Manual on Harmful Marine Microalgae. UNESCO Publishing, Paris, pp. 465–509.

Fuad M.A.M., Mohammad-Noor N., Jalal A.K.C. & Kamaruzzaman B.Y. (2015) Pure cultivation and morphological studies of four *Chaetoceros* taxa from the coastal waters of Pahang, Malaysia. *Sains Malaysiana*, 44: 947–955.

Furnas M.J. (1990) In situ growth rates of marine phytoplankton: approaches to measurement, community and species growth rates. *Journal of Plankton Research*, 12: 1117−1151.

Gačić M., Civitarese G., Miserochhi S., Cardin V., Crise A. & Mauri E. (2002) The open-ocean convection in the Southern Adriatic: a controlling mechanism of the spring phytoplankton bloom. *Continental Shelf Research*, 22: 1897−1908.

Ganley A.R.D. & Kobayashi T. (2007) Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *GenomeResearch*, 17: 184–191.

Gargas A., DePriest P.T. & Taylor J.W. (1995) Positions of multiple insertions in SSU rDNA of lichen-forming fungi. *Molecular Biology and Evolution*, 12: 208–218.

Garrison D.L. (1981) Monterey Bay phytoplankton. 2. Resting spore cycles in coastal diatom populations. *Journal of Plankton Research*, 3: 137–156.

Gilbert W. (1978) Why genes in pieces. *Nature*, 271: 501−501.

Giuffrè G & Ragusa S. (1988) The morphology of *Chaetoceros rostratus* Lauder (Bacillariophyceae) using light & electron microscopy. *Botanica Marina*, 31: 503−510.

Godhe A., Asplund M.E., Härnström K., Saravanan V., Tyagi A. & Karunasagar I. (2008) Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. *Applied Environmental Microbiology*, 74: 7174–7182.

Godrijan J., Marić D., Imešek M., Janeković I., Schweikert M. & Pfannkuchen M. (2012) Diversity, occurrence, & habitats of the diatom genus *Bacteriastrum* (Bacillariophyta) in the northern Adriatic Sea, with the description of *B. jadranum* sp. nov. *Botanica Marina*, 55: 415−426.

Golden B.L. (2008) Ribozymes and RNA catalysis. Cambridge University Press, Cambridge, pp. 178–200.

Gouy M. Guindon S. & Gascuel O. (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment & phylogenetic tree building. *Molecular Biology and Evolution*, 27: 221−224.

Graham J.M., Kent A.D., Lauster G.H., Yannarell A.C., Graham L.E. & Triplett E.W. (2004) Seasonal dynamics of phytoplankton and planktonic protozoan communities in a northern temperate humic lake: diversity in a dinoflagellate dominated system. *Microbial Ecology*, 48: 528–540.

Graham L.E. & Wilcox L. W. (2000) Algae. Prentice-Hall, London. UK.

Gran H.H. (1897) Botany. Protophyta: Diatomaceæ, Silicoflagellata and Cilioflagellata. Den Norske Nordhavs-Expedition 1876–1878. Christiania, Norway. pp. 1–36.

Gran H.H. (1912) The Depths of the Ocean. Macmillan, London, pp. 307–386.

Grube M., Gargas A. & DePriest P.T. (1996) A small insertion in the SSU rDNA of the lichen fungus *Arthonia lapidicola* is a degenerate group-I intron. *Current Genetics*, 29: 582–586.

Guillard R.R.L. & Ryther J.H. (1962) Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt and *Detonula confervacea* Cleve. *Canadian Journal of Microbiology*, 8: 229–239.

Guillard R.R.L. (1975) Culture of marine invertebrate animals Plenum Press, New York, pp. 29−60.

Guillou L., Bachar D. Audic S., Bass D., Berney C., Bittner L., Boutte C., Burgaud G., de Vargas C., Decelle J., del Campo J., Dolan J. R., Dunthorn M., Edvardsen B., Holzmann M., Kooistra W.H.C.F., Lara E., Le Bescot N., Logares R., Mahé F., Massana R., Montresor M.,Morard R., Not F., Pawlowski J., Probert I., Sauvadet A.L., Siano R., Stoeck T., Vaulot D., Zimmermann P. & Christen R. (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote

Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, 41: D597−D604.

Guiry M.D. (2012) How many species of algae are there? *Journal of Phycology*, 48: 1057−1063.

Guiry M.D. & Guiry G.M. (2016) AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. (http://www.algaebase.org; as on 13 February 2016).

Guthrie C. & Patterson B. (1988) Spliceosomal snRNAs. *Annual Review of Genetics*, 22: 387–419.

Hadziavdic K., Lekang K., Lanzen A., Jonassen I., Thompson E.M. & Troedsson C. (2014) Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS One*, 9: e87624.

Hafez M., Iranpour M., Mullineux S.T., Sethuraman J., Wosnitza K.M., Lehn P., Kroeker J., Loewen P.C., Reid J. & Hausner G. (2012) Identification of group I introns within the SSU rDNA gene in species of *Ceratocystiopsis* and related taxa. *Fungal biology*, 116: 98−111.

Hall T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor & analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41: 95−98.

Hamsher S.E., Evans K.M., Mann D.G., Poulíčková A. & Saunders G.W. (2011) Barcoding diatoms: exploring alternatives to COI-5P. *Protist*, 162: 405–422.

Hargraves P.E. & French F.W. (1983) Survival strategies of the algae. Cambridge University Press, Cambridge, pp. 49−68.

Hargraves P.E. (1972) Studies on marine plankton diatoms. I. *Chaetoceros diadema* (Ehr.) Gran: life cycle, structural morphology, & regional distribution. *Phycologia*, 11: 247−257.

Hargraves P.E. (1979) Studies on the marine plankton diatoms IV. Morphology of *Chaetoceros* resting spores. *Nova Hedwigia, Beiheft*, 64: 99−120.

Hasle G.R. & Syvertsen E.E. (1997) Identifying Marine Diatoms & Dinoflagellates. Academic Press, San Diego, pp. 5−385.

Haugen P., Simon D.M. & Bhattacharya D. (2005) The natural history of group I introns. *Trends in Genetics*, 21: 111–119.

Hebert P.D.N., Cywinska A., Ball S.L. & deWaard J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, 270: 313−321

Hendey N. I. (1964) An introductory account of the smaller algae of British coastal waters. V. Bacillariophyceae (Diatoms), HMSO, London.

Hernández-Becerril D.U. (1991) Note on the morphology of *Chaetoceros didymus* & *C. protuberans*, with some considerations on their taxonomy. *Diatom Research*, 6: 289−297.

Hernández-Becerril D.U. (1992) Observations on two closely related species, *Chaetoceros tetrastichon* and *C. dadayi* (Bacillariophyceae). *Nordic Journal of Botany*, 12: 365–371.

Hernández-Becerril D.U. (1993a) Note on the morphology of two planktonic diatoms: *Chaetoceros bacteriastroides* & *C. seychellarus*, with comments on their taxonomy & distribution. *Botanical Journal of the Linnean Society*, 111: 117−128.

Hernández-Becerril D.U. (1993b) Study of the morphology & distribution of two planktonic diatoms: *Chaetoceros paradoxus* & *Ch. filiferus* (Bacillariophyceae) *Cryptogamic Botany*, 3: 169−175.

Hernández-Becerril D.U. (1996) A morphological study of *Chaetoceros* species (Bacillariophyta) from the plankton of the Pacific Ocean of Mexico. *Bulletin of the Natural History Museum of London (Botany)*, 26: 1−73.

Hernández-Becerril D.U. & Granados C.F. (1998) Species of the diatom genus *Chaetoceros* (Bacillariophyceae) in the plankton from the Southern Gulf of Mexico. *Botanica Marina*, 41: 505−519.

Hernández-Becerril D.U., Vilicic D., Bosak S. & D jakovac T. (2010) Morphology & ecology of the diatom *Chaetoceros vixvisibilis* (Chaetocerotales, Bacillariophyceae) from the Adriatic Sea. *Journal of Plankton Research*, 32: 1513−1525.

Hill W.R. (1996) Algal Ecology: Freshwater Benthic Ecosystems. Academic Press, San Diego, pp. 121–148.

Hillebrand H., Dürselen C.D., Kirschtel D., Pollingher U. & Zohary T. (1999) Biovolume calculation for pelagic and benthic microalgae. *Journal of Phycology*, 35: 403–424.

Hillis D.M. & Dixon M.T. (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *Quarterly Review of Biology*, 66: 411–453.

Hood R.R., Wiggert J.D. & Naqvi S.W.A. (2009) Indian Ocean research: opportunities and challenges. *Geophysical Monograph Series*, 409–429.

Hu S.K., Liu Z., Lie A.A., Countway P.D., Kim D.Y., Jones A.C., Gast R.J., Cary S.C., Sherr E.B., Sherr B.F. & Caron D.A. (2015) Estimating protistan diversity using high-throughput sequencing. *Journal of Eukaryotic Microbiology*, 62: 688–693.

Huber J.A., Welch M.D.B., Morrison H.G., Huse S.M., Neal P.R., Butterfield D.A. & Sogin M.L. (2007) Microbial population structures in the deep marine biosphere. *Science*, 318: 97–100.

Huse S.M., Welch D.M., Morrison H.G. & Sogin M.L. (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology*, 12: 1889–1898.

Huseby S., Degerlund M., Zingone A. & Hansen E. (2012) Metabolic fingerprinting reveals differences between northern and southern strains of the cryptic diatom *Chaetoceros socialis*. *European Journal of Phycology*, 47: 480–489.

Hustedt F. (1930) Die Kieselalgen Deutschlands, Osterrreichs und der Schweiz under Beruecksichtigung der uebrigen Lander Europas sowie der angrenzenden Meeresgebiete. 1. Teil. Akademische Verlaggesellschaft, Leipzig.

Ishii K.I., Iwataki M., Matsuoka K. & Imai I. (2011) Proposal of identification criteria for resting spores of *Chaetoceros* species (Bacillariophyceae) from a temperate coastal sea. *Phycologia*, 50: 351−362.

Jensen K. & Moestrup Ø. (1998) The genus *Chaetoceros* (Bacillariophyceae) in inner Danish coastal waters. *Opera Botanica*, 133: 1−68.

Ji B. & Nielsen J. (2015) From next-generation sequencing to systematic modeling of the gut microbiome. *Frontiers in Genetics*, 6: 219.

Johansen S., Elde M., Vader A., Haugen P., Haugli K. & Haugli F. (1997) In vivo mobility of a group I twintron in nuclear ribosomal DNA of the myxomycete *Didymium iridis*. *Molecular Microbiology*, 24: 737−745.

Jolivet A., Chauvaud L., Huchette S., Legoff C., Thébault J., Nasreddine K., Schöne B.R. & Clavier J. (2015) The ormer (*Haliotis tuberculata*): A new, promising paleoclimatic tool. *Palaeogeography*, *Palaeoclimatology*, *Palaeoecology*, 427: 32−40.

Kaczmarska I., Reid C. & Moniz M. (2007) Diatom taxonomy: morphology, molecules and barcodes. *Proceedings of the 1st Central-European Diatom Meeting 2007*, Berlin-Dahlem. pp. 69 72.

Kammerlander B., Breiner H.W., Filker S., Sommaruga R., Sonntag B. & Stoeck T. (2015) High diversity of protistan plankton communities in remote high mountain lakes in the European Alps and the Himalayan mountains. *FEMS Microbiology Ecology*, 91: fiv010.

Karsten G. (1907) Das Indische Phytoplankton nach dem Material der deutschen Tiefsee–Expedition 1898–1899. Wissenschaftliche Ergebnisse der Deutschen Tiefsee–*Expedition auf dem Dampfer 'Valdivia'* 1898–1899 II 2: 221–548

Kartzinel T.R. & Pringle R.M (2015) Molecular detection of invertebrate prey in vertebrate diets: trophic ecology of Caribbean island lizards. *Molecular Ecology Resources*, 15: 903–914.

Katho K., Misawa K., Kuma K.I. & Miyata T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30: 3059−3066.

Khan H. & Archibald J.M. (2008) Lateral transfer of introns in the cryptophyte plastid genome. *Nucleic Acids Research*, 36: 3043–3053.

Ki J.S., Chang K.B., Roh H.J., Lee B.Y., Yoon J.Y. & Jang G.Y. (2007) Direct DNA isolation from solid biological sources without pretreatments with proteinase-K and/or homogenization through automated DNA extraction. *Journal of Bioscience and Bioengineering*, 103: 242−246.

Ki J.S. (2012) Hypervariable regions (V1–V9) of the dinoflagellate 18S rRNA using a large data-set for marker considerations. *Journal of Applied Phycology*, 24: 1035−1043.

Kim S.H. & Cech T.R. (1987) Three-dimensional model of the active site of the self-splicing rRNA precursor of *Tetrahymena*. *Proceedings of the National Academy of Sciences USA*, 84: 8788–8792.

Kooistra W.H.C.F, DeStefano M., Mann D.G. & Medlin L.K. (2003). The phylogeny of the diatoms. In Silicon Biomineralization, Springer, Berlin, Heidelberg, pp. 59-97.

Kooistra W.H.C.F., Sarno D., Balzano S., Gu H., Andersen R.A. & Zingone A. (2008) Global diversity and biogeography of *Skeletonema* species (Bacillariophyta). *Protist*, 159: 177−193.

Kooistra W.H.C.F., Sarno D., Hernández-Becerril D.H., Assmy P., Di Prisco C. & Montresor M. (2010) Comparative molecular and morphological phylogenetic analyses of taxa in the Chaetocerotaceae (Bacillariophyta). *Phycologia*, 49: 471−500.

Kownacka J., Edler L., Gromisz S., Łotocka M., Olenina I., Ostrowska M. & Piwosz K. (2013) Non-indigenous species *Chaetoceros* cf. *lorenzianus* Grunow 1863 - A new, predominant component of autumn phytoplankton in the southern Baltic Sea. *Estuarine, Coastal and Shelf Science*, 119: 101–111.

Kozich J.J., Westcott S.L., Baxter N.T., Highlander S.K. & Schloss P.D. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing

amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*, 79: 5112–5120.

Kühn S.F., Klein G., Halliger H., Hargraves P. & Medlin L.K. (2006) A new diatom, *Mediopyxis helysia* gen. nov. & sp. nov. (Mediophyceae) from the North Sea & the Gulf of Maine as determined from morphological & phylogenetic characteristics. *Nova Hedwigia*, 130: 307−324.

Kuwata A., Hama T. & Takahashi M. (1993) Ecophysiological characterization of two life forms, resting spores & resting cells, of a marine planktonic diatom, *Chaetoceros pseudocurvisetus*, formed under nutrient depletion. *Marine Ecology Progress Series*, 102: 245−255.

Lambowitz A.M. & Belfort M. (1993) Introns as mobile genetic elements. *Annual Review of Biochemistry*, 62: 587–622.

Lange C.B. & Escribano R. (2012) Long-term ocean observations off Central-Southern Chile. Observing the Oceans for Science and Society-Climate, *Ecosystems*. (Ocean_observing_article_COPAS.pdf)

Lauder H.S. (1864) On new diatoms. *Quarterly Journal of Microscopical Science*, 4: 6−8.

Lee S.D. & Lee J.H. (2011) Morphology and taxonomy of the planktonic diatom *Chaetoceros* species (Bacillariophyceae) with special intercalary setae in Korean coastal waters. *Algae*, 26: 153–165.

Lee S.D. & Lee J.H. (2014) *Chaetoceros atlanticus* var. *koreanus* var. nov. (Bacillariophyta) from the East Sea of Korea and closely related taxa: insight into the ultrastructure of the setae. *Diatom Research*, 29: 465−477.

Lee S.D., Park J.S. Yun S.M. & Lee J.H. (2014a) Critical criteria for identification of the genus *Chaetoceros* (Bacillariophyta) based on setae ultrastructure. I. Subgenus *Chaetoceros*. *Phycologia*, 53: 174−187.

Lee S.D., Hyoung M.J. & Lee J.H. (2014b) Critical criteria for identification of the genus *Chaetoceros* (Bacillariophyta) based on setae ultrastructure. II. Subgenus *Hyalochaete*. *Phycologia*, 53: 614−638.

Lefebvre K.A. & Robertson A. (2010) Domoic acid & human exposure risks: a Review. *Toxicon*, 56: 218−230.

Ligowski R., Jordan R.W. & Assmy P. (2012) Morphological adaptation of a planktonic diatom to growth in Antarctic sea ice. *Marine biology*, 159: 817−827.

Li Y., Lundholm N. & Moestrup Ø. (2013) *Chaetoceros rotosporus* sp. nov. (Bacillariophyceae), a species with unusual resting spore formation. *Phycologia*, 52: 600−608.

Li Y., Zhu S., Lundholm N. & Lü S. (2015) Morphology and molecular phylogeny of *Chaetoceros dayaensis* sp. nov. (Bacillariophyceae), characterized by two 90° rotations of the resting spore during maturation. *Journal of Phycology*, 51: 469–479.

Li Z. & Zhang Y. (2005) Predicting the secondary structures and tertiary interactions of 211 group I introns in IE subgroup. *Nucleic Acids Research*, 33: 2118–2128.

Liao D. (1999) Concerted evolution: molecular mechanism and biological implications. *American Journal of Human Genetics*, 64: 24–30.

Lim H.C., Leaw C.P., Su S. N.P., Teng S.T., Usup G., Mohammad-Noor N., Lundholm N., Kotaki Y. & Lim P.T. (2012) Morphology and molecular characterization of *Pseudo-nitzschia* (Bacillariophyceae) from Malaysian Borneo, including the new species *Pseudo-nitzschia circumpora* sp. nov. *Journal of Phycology*, 48: 1232–1247.

Lindeque P.K., Parry H.E., Harmer R.A., Somerfield P.J. & Atkinson A. (2013) Next Generation Sequencing Reveals the Hidden Diversity of Zooplankton Assemblages. *PLoS One*, 8(11): e81327.

Ludwig W. & Klenk H.P. (2001) Bergey's manual of systematic bacteriology (2nd ed.) Springer, New York, pp. 49−65.

Lundholm N., Hasle G.R., Fryxell G.A. & Hargraves P.E. (2002) Morphology, phylogeny and taxonomy of species within the *Pseudo-nitzschia americana* complex (Bacillariophyceae) with descriptions of two new species, *Pseudo-nitzschia brasiliana* and *Pseudo-nitzschia linea*. *Phycologia*, 41: 480−497.

Lundholm N., Moestrup Ø., Kotaki Y., Hoef-Emden K., Scholin C. & Miller P. (2006) Inter- and intraspecific variation of the *Pseudo-nitzschia delicatissima* complex (Bacillariophyceae) illustrated by rRNA probes, morphological data and phylogenetic analyses. *Journal of Phycology*, 42: 464−481.

Lundholm N., Bates S.S., Baugh K.A., Bill B.D., Connell L.B., Léger C. & Trainer V.L. (2012). Cryptic and pseudo-cryptic diversity in diatoms - with descriptions of *Pseudo-nitzschia hasleana* sp. nov. and *P. fryxelliana* sp. nov. I. *Journal of Phycology*, 48: 436–454.

Lykke-Andersen J., Aagaard C., Semionenkov M. & Garrett R.A. (1997) Archaeal introns: splicing, intercellular mobility and evolution. *Trends in Biochemical Sciences*, 22: 326−331.

Lyons P.P., Turnbull J.F., Dawson K.A. & Crumlish M. (2015) Exploring the microbial diversity of the distal intestinal lumen and mucosa of farmed rainbow trout *Oncorhynchus mykiss* (Walbaum) using next generation sequencing (NGS). *Aquaculture Research*, 1: 15.

MacDonald J.D. (1869) On the structure of the diatomaceous frustule, & its genetic cycle. *Annals & Magazine of Natural History*, 4: 1–8.

Machida R.J. & Knowlton N. (2012) PCR Primers for Metazoan Nuclear 18S and 28S Ribosomal DNA Sequences. *PLoS One*, 7: e46180.

Malone T.C., Conley D.J., Fisher T.R., Glibert P.M., Harding L.W. & Sellner K.G. (1996) Scales of nutrient-limited phytoplankton productivity in Chesapeake Bay. *Estuaries*, 19: 371–385.

Malviya S., Scalco E., Audic S., Vincent F., Veluchamy A., Poulaind J., Winckerd P., Iudicone D., de Vargas C., Bittner L., Zingone A. & Bowler C. (2016) Insights into global diatom distribution and diversity in the world's ocean. *Proceedings of the National Academy of Sciences of the USA*, 113: E1516−E1525.

Mangin L. (1910) Sur quelques algues nouvelles ou peu connues du phytoplancton de l'Atlantique. *Bulletin de la Société Botanique de France*, 57: 344–350.

Mann D.G. (1999) The species concept in diatoms. *Phycologia*, 38: 437–495.

Mann D. & Droop S. (1996) Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia*, 336: 19–32.

Mann D.G & Vanormelingen P. (2013) An inordinate fondness? The number, distributions, and origins of diatom species. *Journal of Eukaryotic Microbiology*, 60: 414–420.

Mann K.H. & Lazier J.R.N. (2006) Dynamics of Marine Ecosystems: Biological-Physical Interactions in the Oceans. Blackwell Publishing Ltd. Oxford

Manzari C., Fosso B., Marzano M., Annese A., Caprioli R., D'Erchia A.M., Gissi C., Intranuovo M., Picardi E., Santamaria M., Scorrano S., Sgaramella G., Stabili L., Piraino S. & Pesole G. (2015) The influence of invasive jellyfish blooms on the aquatic microbiome in a coastal lagoon (Varano, SE Italy) detected by an Illumina-based deep sequencing strategy. *Biological Invasions*, 17: 923–940.

Mardis E.R. (2008) Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*, 9: 387–402.

Marino D., Giuffré G., Montresor M. & Zingone A. (1991) An electron microscope investigation on *Chaetoceros minimus* (Levander) comb. nov. & new observations on *Chaetoceros throndsenii* (Marino, Montresor & Zingone) comb. nov. *Diatom Research*, 6: 317−326.

Martin-Jézéquel V., Hildebrand M. & Brzezinski M.A. (2000) Silicon metabolism in diatoms: implications for growth. *Journal of Phycology*, 36: 821−840.

Massana R., Balague V., Guillou L. & Pedros-Alio C. (2004) Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches, *FEMS Microbiology Ecology*, 50: 231–243.

Massana R., Pernice M., Bunge J.A. & del Campo J. (2011) Sequence diversity and novelty of natural assemblages of picoeukaryotes from the Indian Ocean. *The ISME Journal*, 5: 184–195.

Massana R., del Campo J., Sieracki M.E., Audic S. & Logares R. (2014) Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *The ISME Journal*, 8: 854–866.

Massana R., Gobet A., Audic S., Bass D., Bittner L., Boutte C., Chambouvet A., Christen R., Claverie J., Decelle J., Dolan J.R., Dunthorn M., Edvardsen B., Forn I., Forster D., Guillou L., Jaillon O., Kooistra W.H.C.F., Logares R., Mahé F., Not F., Ogata H., Pawlowski J., Pernice M.C., Probert I., Romac S., Richards T., Santini S., Shalchian-Tabrizi K., Siano R., Simon N., Stoeck T., Vaulot D., Zingone A. & de Vargas C. (2015) Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environmental Microbiology*, 17: 4035–4049.

Mavridou A., Cannone J. & Typas M.A. (2000) Identification of group-I introns at three different positions within the 28S rDNA gene of the entomopathogenic fungus *Metarhizium anisopliae* var. *anisopliae*. *Fungal Genetics and Biology*, 31: 79–90.

McQuoid M.R. & Hobson L.A. (1996) Diatom resting stages. *Journal of Phycology*, 32: 889–902.

Medlin L., Elwood H.J., Stickel S. & Sogin M.L. (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene*, 71: 491–499.

Medlin L.K., Elwood H.J., Stickel S. & Sogin M.L. (1991) Morphological and genetic variation within the diatom *Skeletonema costatum* (Bacillariophyta): evidence for a new species, *Skeletonema pseudocostatum*. *Journal of Phycology*, 27: 514–524.

Medlin, L. K., Lange, M., Wellbrock, U., Donner G., Elbrächter M., Hummert C. & Luckas B. (1998) Sequence comparison links toxic European isolates of *Alexandrium tamarense* from the Orkney Islands to toxic north American stocks. *European Journal of Protistology*, 34: 329–335.

Medlin L. & Kaczmarska I. (2004) Evolution of the diatoms: V. Morphological and cytological support for the major clades and a taxonomic revision. *Phycologia*, 43: 245–270.

Menden-Deuer S. & Lessard E.J. (2000) Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton. *Limnology and Oceanography*, 45: 569–579.

Michel F. & Lang B.F. (1985) Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. *Nature*, 316: 641–643.

Michel F., Hanna M., Green R., Bartel D.P. & Szostak J.W. (1989) The guanosine binding site of the *Tetrahymena* ribozyme. *Nature*, 342: 391−395.

Michel F. & Westhof E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *Journal of Molecular Biology*, 216: 585−610.

Mitra S., Laederach A., Golden B.L., Altman R.B. & Brenowitz M. (2011) RNA molecules with conserved catalytic cores but variable peripheries fold along unique energetically optimized pathways. *RNA*, 17: 1589–1603.

Moniz M.B. & Kaczmarska I. (2010) Barcoding of diatoms: nuclear encoded ITS revisited. *Protist*, 161: 7–34.

Montero P., Daneri G., Cuevas L.A., González H.E., Jacob B., Lizárraga L. & Menschel E. (2007) Productivity cycles in the coastal upwelling area off Concepción: the importance of diatoms and bacterioplankton in the organic carbon flux. *Progress in Oceanography*, 75: 518–530.

Montresor M., Sgrosso S., Procaccini G. & Kooistra W.H.C.F. (2003). Intraspecific diversity in *Scrippsiella trochoidea* (Dinophyceae): evidence for cryptic species. *Phycologia*, 42: 56−70.

Montresor M., Di Prisco C., Sarno D. Margiotta F. & Zingone A. (2013) Diversity & germination patterns of diatom resting stages at a coastal Mediterranean site. *Marine Ecology Progress Series*, 484: 79−95.

Morales C.E. & Anabalón V. (2012) Phytoplankton biomass and microbial abundances during the spring upwelling season in the coastal area off Concepción, central-southern Chile: Variability around a time series station. *Progress in Oceanography*, 92–95: 81–91.

Müller K.M., Cannone J.J., Gutell R R. & Sheath R.G. (2001) A structural and phylogenetic analysis of the group ICI introns in the order Bangiales (Rhodophyta). *Molecular Biology and Evolution*, 18: 1654–1667.

Nanjappa D., Kooistra W.H.C.F. & Zingone A. (2013) A reappraisal of the genus *Leptocylindrus* (Bacillariophyta), with the addition of three species and the erection of *Tenuicylindrus* gen. nov. *Journal of Phycology*, 49: 917−936.

Nanjappa D., Audic S., Romac S., Kooistra W.H.C.F. & Zingone A. (2014) Assessment of species diversity and distribution of an ancient diatom lineage using a DNA metabarcoding approach. *PLoS One*, 9(8): e103810.

Narváez D.A., Poulin E., Leiva G., Hernández E., Castilla J.C. & Navarrete S.A., (2004) Seasonal & spatial variation of nearshore hydrographic conditions in central Chile. Continental Shelf Research 24: 279–292.

Navarro J.N. (1982) A survey of the marine diatoms of Pierto Rico III. Suborder Biddulphiineae: Family Chaetocerotaceae. *Botanica Marina*, 25: 305–319.

Neefs J.M., Peer Y.V., Rijk P.D., Chapelle S. & Wachter R.D. (1993) Compilation of small ribosomal subunit RNA structures. *Nucleic Acids Research*, 21: 3025−3049.

Nelles L., Fang B.L., Volckaert G., Vandenberghe A. & Wachter R.D. (1984) Nucleotide sequence of a crustacea 18S ribosomal RNA gene and secondary structure of eukaryotic small subunit ribosomal RNAs. *Nucleic Acids Research*, 12: 8749−8768.

Nickrent D.L. & Sargent M.L. (1991) An overview of the secondary structure of the V4 region of eukaryotic small-subunit ribosomal RNA. *Nucleic Acids Research*, 19: 227−235.

Not F., Siano R., Kooistra W.H.C.F., Simon N., Vaulot D. & Probert I. (2012) Diversity and Ecology of Eukaryotic Marine Phytoplankton. *Advances in Botanical Research*, 64: 1−53.

Officer C.B. & Ryther J.H. (1980). The possible importance of silicon in marine eutrophication. *Marine Ecology Progress Series*, 3: 83–91.

Okuno H. (1962) Electron-microscopical study on fine structures of diatom frustules. XIX. *The Botanical Magazine*. Tokyo 75: 119−126.

Olsen G.J., Lane D.J., Giovannoni S.J., Pace N.R. & Stahl D.A. (1986) Microbial ecology and evolution: a ribosomal RNA approach. *Annual Review of Microbiology*, 40: 337–365.

Orgiazzi A., Dunbara M.B., Panagosa P., de Grootb G.A. & Lemanceau P. (2015) Soil biodiversity and DNA barcodes: opportunities and challenges. *Soil Biology and Biochemistry*, 80: 244–250.

Ostenfeld C.H. (1903) Phytoplankton from the sea around the Faeröes. Botany of the Faeröes Based Upon Danish Investigations Part II. Copenhagen: *Det Nordiske Forlag*. pp. 558–612.

Ostenfeld C.H. (1913). On the distribution of Bacillariales (diatoms) in the plankton of the North European waters according to the international sea investigations, with special relation to the hydrographical conditions. Copenhague: *Bureau du Conseil permanent international pour l'exploration de la mer*. pp. 403–508, plate 54–93.

Pánek T., Táborský P., Pachiadaki M.G., Hroudová M., Vlček Č., Edgcomb V.P. & Čepička I. (2015) Combined culture-based and culture-independent approaches provide insights into diversity of jakobids, an extremely plesiomorphic eukaryotic lineage. *Frontiers in Microbiology*, 6: 1288.

Padgett R.A., Grabowski P.J., Konarska M.M., Seiler S. & Sharp P.A. (1986) Splicing of messenger RNA precursors. *Annual Review of Biochemistry*, 55: 1119–1150.

Pavillard J. (1916) Recherches sur les diatomées pélagiques du Golfe du Lion. Travail de l'lnstitut de Botanique de l'Université de Montpellier et de la Station Zoologique de Cette. *Serie mixte, memoire*, 5: 7−62.

PavillardJ. (1925) Bacillariales. Report on the Danish Oceanographical Expeditions 1908−1910 to the Mediterranean & adjacent Seas, vol. II, *Biology*, 9: 1−72.

Pawlowski J., Christen R., Lecroq B., Bachar D., Shahbazkia H.R., Amaral-Zettler L.A. & Guillou L. (2011) Eukaryotic richness in the abyss: insights from pyrotag sequencing. *PLoS One*, 6:e18169.

Pereira F., Carneiro J. & Amorim A. (2008) Identification of species with DNA−based technology: current progress and challenges. *Recent Patents on DNA & Gene Sequences*, 2: 187–200.

Pfitzer E. (1869) Über den Bau und die Zellteilung der Diatomeen. *Botanische Zeitung*, 27: 774−776.

Pielou E.C. (2008) After the Ice Age: The return of life to glaciated North America. University of Chicago Press, Chicago.

Piredda R., Tomasino M.P., D'Erchia A.M., Manzari C., Pesole G., Montresor M., Kooistra W.H.C.F., Sarno D. & Zingone A. (2017) Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean LTER site. *FEMS Microbiology*, 93 (1): fiw200.

Podar M., Dib-Hajj S. & Perlman P.S. (1995) A UV-induced, Mg(2+)-dependent crosslink traps an active form of domain 3 of a self-splicing group II intron. *RNA*, 1: 828–840.

Pondaven P., Gallinari M., Chollet S., Bucciarelli E., Sarthou G., Schultes S., & Jean F. (2007) Grazing-induced changes in cell wall silicification in a marine diatom. Protist, 158: 21–28.

Price M.N., Dehal P.S. & Arkin A.P. (2009) FastTree: Computing Large Minimum-Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution*, 26: 1641–1650.

Price M.N., Dehal P.S. & Arkin A.P. (2010) FastTree2 - Approximately maximum-likelihood trees for large alignments. *PLoS One*, 5: e9490.

Prokopowich C.D., Gregory T.R. & Crease T.J. (2003) The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, 46: 48–50.

Proschkina-Lavrenko A.I. (1963) Diatomovye Vodorosli Plankton Azovskogo Morya. Akademija Nauk SSSR, Botaniseskie Institute, Moscow–Leningrad.

Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Peplies J. & Glöckner F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing & web-based tools. *Nucleic Acids Research*, 41: D590−D596.

Quijano-Scheggia S.I., Garcés E., Lundholm N., Moestrup Ø., Andree K. & Camp J. (2009) Morphology, physiology, molecular phylogeny and sexual compatibility of the cryptic *Pseudo-nitzschia delicatissima* complex (Bacillariophyta), including the description of *P. arenysensis* sp. nov. *Phycologia*, 48: 492−509.

Reyes A., Semenkovich N.P., Whiteson K., Rohwer F. & Gordon J.I. (2012) Going viral: next-generation sequencing applied to human gut phage populations. *Nature Reviews Microbiology*, 10: 607–617.

Reynolds C.S. & Walsby A.E. (1975) Water-blooms. *Biological reviews*. 50: 437−481.

Ribera d'Alcalà M., Conversano F., Corato F., Licandro P., Mangoni O., Marino D., Mazzocchi M.G., Modigh M., Montresor M., Nardella M. & Saggiomo V. (2004). Seasonal patterns in plankton communities in a pluriannual time series at a coastal Mediterranean site (Gulf of Naples): an attempt to discern recurrences and trends. *Scientia Marina*, 68: 65−83.

Rines J.E.B. & Hargraves P.E. (1987) The seasonal distribution of the marine genus *Chaetoceros* Ehr. in Narragansett Bay, Rhode Island (1981–1982). *Journal of Plankton Research*, 9: 917−933.

Rines J.E.B & Hargraves P.E. (1988) The *Chaetoceros* Ehrenberg (Bacillariophyceae) flora of Narragansett Bay, Rhode Island, USA. *Bibliotheca Phycologica*, 79: 1−196.

Rines J.E.B. & Hargraves P.E. (1990) Morphology & taxonomy of *Chaetoceros compressus* Lauder var. *hirtisetus* var. nova, with preliminary consideration of closely related taxa. *Diatom Research*, 5: 113−127.

Rines J.E.B. (1999) Morphology & taxonomy of *Chaetoceros contortus* Schütt 1895, with preliminary observations on *Chaetoceros compressus* Lauder 1864 (Subgenus *Hyalochaete*, Section *Compressa*). *Botanica Marina*, 42: 539−551.

Rines J.E.B. & Theriot E.C. (2003) Systematics of Chaetocerotaceae (Bacillariophyceae) I. A phylogenetic analysis of the family. *Phycological Research,* 51: 83−98.

Rogers S.O., Yan Z.H., Shinohara M., Lobuglio K.F. & Wang C.J.K. (1993) Messenger RNA intron in the nuclear 18S ribosomal RNA gene of deuteromycetes. *Current Genetics*, 23:338–342.

Ronquist F. & Huelsenbeck J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19: 1572–1574.

Round F.E., Crawford R.M. & Mann D.G. (1990) The Diatoms: Biology & Morphology of the Genera. Cambridge University Press, Cambridge.

Rousseau V., Chretiennot-Dinet M.J., Whipple S., Verity P. & Jacobsen A. (2007) The life cycle of *Phaeocystis*: state of knowledge and presumptive role in ecology. *Biogeochemistry*, 83: 29–47.

Ruiz M.J.L., Soto P.J., Zamudio M.E., Hernández-Becerril D.U. & Licea D.S. (1993) Morphology and taxonomy of *Chaetoceros diversus* (Bacillariophyceae) based on material from the southern Gulf of Mexico. *Diatom Research*, 8: 419–428.

Sackett O., Petrou K., Reedy B., De Grazia A., Hill R., Doblin M., Beardall J. Ralph P. & Heraud P. (2013) Phenotypic plasticity of southern ocean diatoms: key to success in the sea ice habitat? *PLoS One*, 8: e81185.

Saldanha R., Mohr G., Belfort M. & Lambowitz A.M. (1993) Group I and group II introns. *Federation of American Societies for Experimental Biology*, 7: 15–24.

Samanta B. & Bhadury P. (2016) A comprehensive framework for functional diversity patterns of marine chromophytic phytoplankton using rbcL phylogeny. *Scientific Reports*, 6: 20783.

Sancetta C. & Calvert S.E. (1988) The annual cycle of sedimentation in Saanich inlet, British Columbia: implications for the interpretation of diatom fossil assemblages. *Deep Sea Research Part A. Oceanographic Research Papers*, 35: 71−90.

Sánchez Castillo P.M., Leon U.M.A. & Round F.E. (1992) Estudio de *Chaetoceros wighamii* Brightwell: un taxon mal interpretado. *Diatom Research*, 7: 127–136.

Sanger F., Nicklen S. & Coulson A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA*, 74: 5463–5467.

Sar E.A., Hernández-Becerril D.U. & Sunesen I. (2002) A morphological study of *Chaetoceros tenuissimus* Meunier, a little-known planktonic diatom, with a discussion of the section *Simplicia*, subgenus *Hyalochaete*. *Diatom Research*, 17: 327−335.

Sarno D., Zingone A. & Marino D. (1997) *Bacteriastrum parallelum* sp. nov., a new diatom from the Gulf of Naples, & new observations on *B. furcatum* (Chaetocerotaceae, Bacillariophyta). *Phycologia*, 36: 257−266.

Sarno D., Kooistra W.C.H.F., Medlin L.K., Percopo I. & Zingone A. (2005) Diversity in the genus *Skeletonema* (Bacillariophyceae). II. An assessment of the taxonomy of *S. costatum*-like species, with the description of four new species. *Journal of Phycology*, 41: 151–176.

Sarno D., Kooistra W.H.C.F, Balzano S., Hargraves P.E. & Zingone A. (2007) Diversity in the genus *Skeletonema* (Bacillariophyceae): III. Phylogenetic position and morphological variability of *Skeletonema costatum* and *Skeletonema grevillei*, with the description of *Skeletonema ardens* sp. nov. *Journal of Phycology*, 43: 156-170.

Savidge G., Boyd P., Pomroy A., Harbour D. & Joint I. (1995) Phytoplankton production & biomass estimates in the north-east Atlantic Ocean, May-June 1990. *Deep Sea Research* I, 42: 599−617.

Savin M.C., Martin J., LeGresley M., Giewat M. & Rooney-Varga J. (2004) Plankton diversity in the Bay of Fundy as measured by morphological & molecular methods. *Microbial Ecology*, 48: 51−63.

Schloss P.D., Westcott S.L., Ryabin T., Hall J.R., Hartmann M., Hollister E.B., Lesniewski R.A., Oakley B.B., Parks D.H., Robinson C.J. & Sahl J.W. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, 75: 7537–7541.

Schmidt R., Mäusbacher R. & Müller J. (1990) Holocene diatom flora and stratigraphy from sediment cores of two Antarctic lakes (King George Island). *Journal of Paleolimnology*, 3: 55–74.

Scholin C.A., Herzog M., Sogin M. & Anderson D.M. (1994) Identification of group- and strain-specific genetic markers for globally distributed *Alexandrium* (Dinophyceae). II. Sequence analysis of a fragment of the LSU rRNA gene. *Journal of Phycology*, 30: 999–1011.

Schütt F. (1895). Arten von *Chaetoceras* und *Peragallia*. Ein Beitrag zur Hochseeflora. *Berichte der Deutsche Botanisch Gesellschaft*, 13: 35–50.

Scotto di Carlo B., Tomas C.R., Ianora A. Marino J, Mazzocchi M.G., Modigh M., Montresor M., Petrillo L., Ribera d'Alcala M., Saggiomo V. & Zingone A. (1985) Uno studio integrato de U'ecosistema pelagico costiero del golfo di Napoli. *Nova Thalassia*, 7: 99−128.

Shadbolt G. (1854) A short description of some new forms of Diatomaceae from Port Natal. *Transactions of the Microscopical Society of London*, 2: 13−18.

Shevchenko O.G., Orlova T.Y. & Hernández-Becerril D.U. (2006) The genus *Chaetoceros* (Bacillariophyta) from Peter the Great Bay, Sea of Japan. *Botanica Marina*, 49: 236−258.

Shokralla A., Spall J.L., Gibson J.F. & Hajibabaei M. (2012) Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, 21: 1794−1805.

Shrestha R.P. & Hildebrand M. (2015) Evidence for a regulatory role of diatom silicon transporters in cellular silicon responses. *Eukaryotic Cell,* 14: 29−40.

Silvestro D. & Michalak I. (2012) raxmlGUI: a graphical front-end for RaxML. *Organisms Diversity & Evolution*, 12: 1–3.

Sims P.A., Mann D.G. & Medlin L.K. (2006) Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia*, 45: 361–402.

Smetacek V. (1999) Diatoms & the ocean carbon cycle. *Protist*, 150: 25−32.

Smol J.P., Battarbee R.W., Davies R.B. & Meriläinen J. (1986) Diatoms & Lake Acidity. Reconstructing pH from siliceous algal remains in lake sediments. Developments in Hydrobiology. Dr W. Junk Publishers, Dordrecht.

Snoeijs P. (1999) Marine and brackish waters. *Acta Phytogeographica Suecica*, 84: 187–212

Sobarzo M., Bravo L., Donoso D., Garcés-Vargas J. & Schneider W. (2007) Coastal upwelling & seasonal cycles that influence the water column over the continental shelf off central Chile. *Progress in Oceanography*, 75: 363−382.

Sogin M.L., Morrison H.G., Huber J.A., Welch D.M., Huse S.M., Neal P.R., Arrieta J.M. & Herndl G.J. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences USA*, 103: 12115–12120.

Sonnenberg R., Nolte A.W. & Tautz D. (2007) An evaluation of LSU rDNA D1–D2 sequences for their use in species identification. *Frontiers in Zoology*, 4: 6.

Sorhannus U., Gasse F., Perasso R. & Tourancheau A.B. (1995) A preliminary phylogeny of diatoms based on 28S ribosomal RNA sequence data. *Phycologia*, 34: 65–73.

Sorhannus U. (2004) Diatom phylogenetics inferred based on direct optimization of nuclear-encoded SSU rRNA sequences. *Cladistics*, 20: 487–497.

Sorhannus U. (2007) A nuclear-encoded small-subunit ribosomal RNA timescale for diatom evolution. *Marine Micropaleontology*, 65: 1–12.

Sournia A., Birrien J.L., Douville J.L., Klein B. & Viollier M. (1987) A daily study of the diatom spring bloom at Roscoff (France) in 1985. I. The spring bloom within the annual cycle. *Estuarine*, *Coastal and Shelf Science*, 25: 355-367.

Spatharis S., Danielidis D.B. & Tsirtsis G. (2007) Recurrent *Pseudo-nitzschia calliantha* (Bacillariophyceae) & *Alexandrium insuetum* (Dinophyceae) winter blooms induced by agricultural runoff. *Harmful Algae*, 6: 811−822.

Stamatakis A., Hoover P. & Rougemont J. (2008) A rapid bootstrap algorithm for the RAxML web-servers. *Systematic Biology*, 75: 758−771.

Stamatakis A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30: 1312–1313.

Stevenson R.J., Bothwell M.L., Lowe R. L. & Thorp J.H. (1996) Algal ecology: *Freshwater benthic ecosystem*. Academic press.

Stickley C.E., Koç N., Brumsack H.J., Jordan R.W. & Suto I. (2008) A siliceous microfossil view of middle Eocene Arctic paleoenvironments: a window of biosilica production & preservation. *Paleoceanography*, 23: 19.

Stoeck T. & Epstein S. (2003) Novel eukaryotic lineages inferred from small-subunit rRNA analyses of oxygen-depleted marine environments. *Applied and Environmental Microbiology*, 69: 2657–2663.

Stoeck T., Bass D., Nebel M., Christen R., Jones M.D., Breiner H.W. & Richards T.A. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, 19: 21–31.

Sugie K. & Kuma K. (2008) Resting spore formation in the marine diatom *Thalassiosira nordenskioeldii* under iron- & nitrogen-limited conditions. *Journal of Plankton Research*, 30: 1245−1255.

Sun J. & Liu D. (2003) Geometric models for calculating cell biovolume and surface area for phytoplankton. *Journal of Plankton Research*, 25: 1331–1346.

Sunesen I., Hernández-Becerril D.U. & Sar E.A. (2008) Marine diatoms from Buenos Aires coastal waters (Argentina). V. Species of the genus *Chaetoceros*. *Revista de biología marina y oceanografía*, 43: 303–326.

Suto I. (2006) The explosive diversification of the diatom genus *Chaetoceros* across the Eocene/Oligocene & Oligocene/Miocene boundaries in the Norwegian Sea. *Marine Micropaleontology*, 58: 259−269.

Sverdrup H.U. (1953) On conditions for the vernal blooming of phytoplankton. *Journal du Conseil*, 18: 287–295.

Swofford D.L. (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Takishita K., Yubuki N., Kakizoe N., Inagaki Y. & Maruyama T. (2007) Diversity of microbial eukaryotes in sediment at a deep-sea methane cold seep: surveys of ribosomal DNA libraries from raw sediment samples & two enrichment cultures. *Extremophiles*, 11: 563–576.

Tamura K., Stecher G., Peterson D., Filipski A. & Kumar S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30: 2725–2729.

Tautz D., Arctander P., Minelli A., Thomas R.H. & Vogler A.P. (2003) A plea for DNA taxonomy. *Trends in Ecology & Evolution*, 18: 70–74.

Theriot E.C., Cannone J.J., Gutell R.R. & Alverson A.J. (2009) The limits of nuclear-encoded SSU rDNA for resolving the diatom phylogeny. *European Journal of Phycology*, 44: 277−290.

Theriot E.C., Ashworth M., Ruck E., Nakov T. & Jansen R.K. (2010) A preliminary multigene phylogeny of the diatoms (Bacillariophyta): challenges for future research. *Plant Ecology & Evolution*, 143: 278−296.

Tocchini-Valentini G.D., Fruscoloni P. & Tocchini-Valentini G.P. (2011) Evolution of introns in the archaeal world. *Proceedings of the National Academy of Sciences USA*, 108: 4782–4787.

Treguer P., Nelson D.M., VanBennekom A.J. & DeMaster D.J. (1995) The silica balance in the world ocean: a reestimate. *Science*, 268: 375−379.

Vader A., Naess J., Haugli K., Haugli F. & Johansen S. (1994) Nucleolar introns from *Physarum flavicomum* contain insertion elements that may explain how mobile group I introns gained their open reading frames. *Nucleic Acids Research*, 22: 4553–4559.

Van De Peer Y. & Wachter R.D. (1997) Evolutionary Relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. *Journal of Molecular Evolution*, 45: 619−630.

Van Dijk E.L., Auger H., Jaszczyszyn Y. & Thermes C. (2014) Ten years of next-generation sequencing technology. *Trends in genetics*, 30: 418−426.

Vanormelingen P, Chepurnov V.A., Mann D.G., Sabbe K. & Vyverman W. (2008) Genetic divergence & reproductive barriers among morphologically heterogeneous sympatric clones of *Eunotia bilunaris sensu lato* (Bacillariophyta). *Protist*, 159: 73−90.

Vicens Q. & Cech T.R. (2006) Atomic level architecture of group I introns revealed. *Trends in Biochemical Sciences*, 31: 41–51.

Visco J.A., Gentil L.A.P., Cordonier A., Esling, P., Pillet, L. & Pawlowski J. (2015) Environmental monitoring: Inferring the diatom index from next generation sequencing data. *Environmental Science and Technology*, 49: 7597−7605.

Vogler A.P. & Monaghan M.T. (2007) Recent advances in DNA taxonomy. *Journal of Zoology and Systematic Evolution Research*, 45: 1−10.

Von Stosch H.A., Theil G. & Kowallik K.V. (1973) Entwicklungsgeschichtliche Untersuchungen an zentrischen Diatomeen. V. Bau und Lebenszyklus von *Chaetocerosdidymum*, mit Beobachtungen ü ber einige andere Arten der Gattung. *Helgolander Wissenschaftliche Meeresuntersuchungen*, 25: 384−445.

Wang Q., Garrity G.M., Tiedje J.M. & Cole J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73: 5261−5267.

Wang J., Wang F., Chu L., Wang H., Zhong Z., Liu Z., Gao J. & Duan H. (2014) High Genetic Diversity and Novelty in Eukaryotic Plankton Assemblages Inhabiting Saline Lakes in the Qaidam Basin. *PLoS One*, 9: e112812.

Wiggert J.D. & Murtugudde R.G. (2007) The sensitivity of the Southwest Monsoon phytoplankton bloom to variations in aeolian iron deposition over the Arabian Sea. *Journal of Geophysical Research*, 112: C5005.

Wieters E.A., Kaplan D.M., Navarrete S.A., Sotomayor A., Largier J., Nielsen K.J. & Veliz F. (2003) Alongshore and temporal variability in chlorophyll a concentration in Chilean near shore waters. *Marine Ecology Progress Series*, 249: 93–105.

Williams D.M. & Kociolek J.P. (2007) Pursuit of a natural classification of diatoms: History, monophyly and the rejection of paraphyletic taxa. *European Journal of Phycology*, 42: 313–319.

Woese C.R. (1987) Bacterial evolution. *Microbiological reviews*, 51: 221–271.

Woodson S.A. (2005) Structure & assembly of group I introns. *Current Opinion in Structural Biology*, 15: 324−330.

Wuyts J., Van de Peer Y. & Wachter R.D. (2001) Distribution of substitution rates & location of insertion sites in the tertiary structure of ribosomal RNA. *Nucleic Acids Research*, 29: 5017–5028.

Yamada T., Tamura K., Aimi T. & Songsri P. (1994) Self-splicing group I introns in eukaryotic viruses. *Nucleic Acids Research*, 22: 2532–2537.

Yang C.Z. & Albright L.J. (1992) Effects of the harmful diatom *Chaetoceros concavicornis* on respiration of rainbow trout *Oncorhynchus mykiss*. *Diseases of Aquatic Organisms*, 14: 105−114.

Yilmaz P., Parfrey L.W., Yarza P., Gerken J., Pruesse E., Quast C., Schweer T., Peplies J., Ludwig W. & Glöckner F.O. (2014) The SILVA & "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research*, 42: D643−D648.

Zhu F., Massana R., Not F., Marie D. & Vaulot D. (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology*, 52: 79–92.

Zimmermann J., Glöckner G., Jahn R., Enke N. & Gemeinholzer B. (2014) Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Molecular Ecology Resources*, 15: 526–542.

Zinger L., Gobet A. & Pommier T. (2012) Two decades of describing the unseen majority of aquatic microbial diversity. *Molecular Ecology*, 21: 1878−1896.

Zingone A., Montresor M. & Marino D. (1990) Summer phytoplankton physiognomy in coastal waters of the Gulf of Naples. *Marine Ecology*, 11: 157−172.

Zingone A., Casotti R., d'Alcalà M.R., Scardi M. & Marino D. (1995) 'St Martin's Summer': the case of an autumn phytoplankton bloom in the Gulf of Naples (Mediterranean Sea). *Journal of Plankton Research*, 17: 575–593.

Zingone A. & Sarno D. (2001) Recurrent patterns in coastal phytoplankton from the Gulf of Naples. *Archivio di Oceanografia e Limnologia*, 22: 113–118.

Zingone A., Dubroca L., Iudicone D., Margiotta F., Corato F., d'Alcalà M.R., Saggiomo V. & Sarno D. (2010) Coastal phytoplankton do not rest in winter. *Estuaries and Coasts*, 33: 342–361.

Allopatry: species occurring in different geographic area.

Aperture: the opening between the valves of the adjacent cells.

Apical axis: the long axis of the valve face, passing through the apices.

Apomorphy: a shared derived character state of a descendent or a novel evolutionary character state.

Autapomorphy: a character state unique to a descendent.

Auxospore: a specialised cell that is the product of gamete fusion in diatoms. It is used to restore the maximum cell size.

Barcode: a fragment of a gene sequence from a standardized portion of the genome to identify species.

Biodiversity: the variety of life on earth. It can be variety within- or between all the species and their interaction with the environment in which they live.

Character: a heritable feature of an organism that is used to distinguish or differentiate a species.

Chimera sequence: a sequence that consists of nucleotide fragments of two individuals.

Cingulum: the series of siliceous bands associated with a valve. The epivalve and its associated epicingulum are called epitheca; the hypovalve and its associated hypocingulum are called hypotheca.

Clade: a group of organisms including a common ancestor and all – and only - its descendants.

Cladogenesis: a process of branching or divergence leading to speciation.

Cladogram: a cladistics diagram used to depict a hypothetical branching sequence of lineages leading to taxa under study.

Concerted evolution: processes that keep a string of DNA copies (multi-gene families) homogenized within a population.

Contigs: an overlapping sequence read, usually a forward and a reverse read of a sequence combined.

Convergence: similarities in unrelated groups of organisms.

Convergent evolution: independent acquisition of a character state with similar function in two lineages.

Cryptic: two or more species that do not interbreed anymore, but that still share identical morphology.

Diplontic: a species in which mitosis happens only in the diploid phase of the life cycle. Following gametogenesis, gametes conjugate to form the diploid F1.

Epilithic: species that grow on the surface of rock or stone.

Epipelic: algae that grow on the surface of sediments or clay or silt.

Epiphytic: algal species that live on the surface of other algae, without harming them.

Epipsammic: alge attached to -or moving through- sand particles.

Epitheca: the larger theca of the frustule, which overlaps over the hypotheca.

Epizoic: algal species growing on the external surface of animals.

Frustule: the siliceous part of the diatom cell wall, which comprises of the epitheca and hypotheca.

Gene: the basic unit of heredity, or the functional and physical unit of inheritance.

Genetic diversity: the variation of the genetic information within and among the individuals of a species, population or a community.

Grade: an artificial taxon. A grade is paraphyletic.

Halocline: a vertical zone in the ocean water column where salinity changes rapidly with depth.

Haplotype: a DNA sequence composed of a particular string of nucleotides. Within the context of HTS contigs I use haplotype as a set of contigs, clustered at the 99% similarity-level.

Heterothallic: species that have sexes in different individuals, i.e. male and female gametophyte are different individuals.

Holotype: a single specimen clearly designated in the original description.

Homology: a shared character in different groups of organisms which is derived from the same structure or trait in their common ancestor.

Homoplasy: a character state with similar function seen in different groups of organisms, which are not derived from a common ancestor, but is the result of convergent evolution or character reversal.

Homothallic: species that have both sexes in the same individual, i.e., the same gametophyte can produce both male and female gametes.

Horizontal gene transfer: a natural process in which DNA sequences of an organism are transferred and incorporated in the genome of other organisms belonging to different species.

Hyaline rim: is a siliceous structure that extends from the marginal ridge, it can be variable in height.

Hypotheca: the smaller theca of the frustule

Ingroup: a group that is investigated in the study.

Internode: is the line, which connects two subsequent speciation events.

Interspecific: occurring or arising between species.

Intraspecific: occurring or arising within a species.

Intron: a fragment of a gene that occupies the DNA sequence and does not code for any protein.

Isogamy: a form of sexual reproduction in which the male and female gametes show similar morphology.

Isotype: is the duplicate of the holotype.

Isovalvate: the two valves of the frustule have similar morphology.

Lineage: a group of organisms that are linked by a continuous line of descent, i.e., parent – offspring connection.

MacDonald-Pfitzer rule: at each mitotic cell division the new valve is formed inside the parental theca, causing the average size of the frustules or valve in a population to decrease slowly.

Meta-barcode haplotypes: are the haplotypes generated for the meta-barcoding approach.

Meta-barcoding: is an approach used for rapid assessment of biodiversity that integrates two technologies, i.e., DNA based identification and high-throughput DNA sequencing.

Neritic: a zone of the ocean waters from the low-tide mark to the edge of the continental shelf at a depth of ca. 200 meters, characterised by stable temperature, low water pressure and plenty of sunlight.

Node: a representation of the speciation event in the tree.

Oligotrophic water: a water body characterised by extremely low nutrient concentrations for nitrogen and phosphorous, the result of which is low productivity.

Oogamy: is a form of sexual reproduction where the male and female gametes are morphologically distinguishable.

Outgroup: a group in a phylogeny, which is supposed sister to the taxon under study.

Parallel evolution: development of similar character states in species or taxonomic groups that are not directly related by evolutionary descent.

Paraphyletic: a group of organisms sharing a last common ancestor, but not all the descendants of that last common ancestor belong to the group.

Parsimony: a rule used to choose a best cladogram from the possible cladograms by implying the least number of changes in character states.

Pelagic: belonging to the water column of the open sea.

Pervalvar axis: the axis of the valve that is perpendicular to the centre of the valve face.

Phylogenetics: a scientific discipline to resolve the phylogenetic relationships between organisms.

Phylogeny: the evolutionary relationships among organisms.

Plesiomorphy: a primitive character state that is ancestral within the taxa under consideration.

Polyphyletic: a group of organisms or taxa, which do not include their last common ancestor.

Polytomy: is the branch point of a tree that has more than two immediate descendants.

Population: a group of specimens that freely interbreed.

Pseudo-cryptic species: species that can be differentiated morphologically once the genetics are known.

Pycnocline: a level in the water column at which water density changes abrubtly.

Reference barcode: a standardized gene fragment that is associated with a species whose morphological and molecular information is validated.

Resting spore: a dormant stage in diatoms, which is characterised by a protective coat of silica, adapted to resist unfavourable conditions.

Resting stage: a dormant cell of an organism, which aids survival of unfavourable conditions. Resting stages resemble vegetative cells.

Rimoportula: also referred as labiate process, is a tube through the valve of a diatom. In the valve face, internally it shows a slit-shaped or lip-shaped opening whereas externally the opening may be simple, round or tube shaped.

Semi-cryptic: somewhat cryptic or enigmatic species.

Sequence: a succession of letters (A, C, G, T or U) that represents the order of nucleotides in nucleic acids (DNA or RNA).

Seta: the siliceous extension of the Chaetocerotacean valve, which can be simple or robust.

Sister groups: two groups that result from splitting of a single lineage.

Special setae: intercalary setae that can be differentiated from the normal setae.

Speciation: an event where a lineage produces two new, separate species, which in turn begin their own unique evolutionary histories.

Species complex: a group of closely related species that are difficult to be separated morphologically.

Species: in the present context, defined as a group of individuals that show reproductive isolation and genetic differentiation.

Spine: pointed siliceous extension on the valve or setae of a vegetative cell or on a resting spore.

Sympatry: two or more species occupying the same geographic area.

Symplesiomorphy: a shared ancestral character state.

Synapomorphy: a shared derived character state.

Taxon: a group of organisms that are named.

Terminal clade: a clade bearing usually short branches and without further marked internal structure, i.e., there are no well-supported clades inside.

Terminal valves: the valves at the termini of the colony. In Chaetoceroataceae, the terminal valve is characterised by setae that do not join with corresponding sibling setae. Terminal valves always have a rimoportula.

Theca: a valve and its associated cingulum.

Thermocline: a distinct transition in a water column at which the temperature changes rapidly; usually the transition between warmer mixed water near the ocean surface and cooler deep water below.

Transapical axis: the short axis of the valve face, passing across the middle of the face, perpendicular to the apical axis.

Valve: the intricately sculptured unit of the diatom frustule, displaying most of the ornamentation used for identification.

Valvocopula: the first girdle band or copula, which is in direct contact with the valve.

Vertical gene transfer: a natural process where the genes of an organism are transferred to the offspring.